National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Services des thèses canadiennes

Ottawa, Canada
K1A 0N4

CANADIAN THESES
THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this film is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de ce microfilm est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

THIS DISSERTATION
HAS BEEN MICROFILMED
EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ
MICROFILMÉE TELLE QUE
NOUS L'AVONS REÇUE

NL-339 (r. 86/06)

Canada

THE UNIVERSITY OF ALBERTA

NUCLEOTIDE SEQUENCE OF THE CARBOXY-TERMINAL PORTION OF A LODGEPOLE PINE ACTIN GENE

by

JOYCE R. KENNY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF FOREST SCIENCE

EDMONTON, ALBERTA

FALL 1986

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-32324-8

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR JOYCE R. KENNY

TITLE OF THESIS NUCLEOTIDE SEQUENCE OF THE CARBOXY-TERMINAL PORTION OF A LODGEPOLE PINE ACTIN GENE

DEGREE FOR WHICH THESIS WAS PRESENTED MASTER OF SCIENCE

YEAR THIS DEGREE GRANTED FALL 1986

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) .....

PERMANENT ADDRESS:

.....1512.....109.....ST.....

.....EDMONTON, ALTA.....

.....T6J.....5W8.....

DATED .....Aug.....19 86

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled NUCLEOTIDE SEQUENCE OF THE CARBOXY-TERMINAL PORTION OF A LODGEPOLE PINE ACTIN GENE submitted by JOYCE R. KENNY in partial fulfilment of the requirements for the degree of MASTER OF SCIENCE.

Bruce P. Dancik
.....

Supervisor

.....

.....

.....

.....

Date.....

ABSTRACT

The nucleotide and amino acid sequence of the carboxy-terminal portion of a lodgepole pine actin gene, PAc1-A, were determined. Comparison of PAc1-A to β-actin genes from soybean, maize, chicken, and yeast has shown that although the actin gene is highly conserved among all these species, consistent differences can be detected to differentiate the actins. The plant actins shared the lowest amount of divergence (or highest homology values) when compared among themselves, and the highest amount of divergence when compared to yeast. It is likely that conservation of actin on a broad scale reflects the importance of actin as a 'housekeeping' gene for carrying out basic cellular functions. On a much smaller scale, differences existing among the various species' actins may have resulted from subtle cellular differences among organisms. Thus actins may be classified according to their different roles.

Although it seems likely that actin belongs to a multigene family, problems presented by the large genome size and contamination of genomic lodgepole pine DNA with plasmids used in the laboratory makes accurate determination of the number of genes difficult. Based on results from *EcoRI*- and *HindIII*-digested pine genomic DNA hybridized first with ^{32}P-labelled PAc1-A, and later washed and re-hybridized with ^{32}P-labelled pBR322, the number of actin genes was estimated to be from four to six.

ACKNOWLEDGEMENTS

I would like to thank my Supervisor Dr. B.P. Dancik for his support throughout the course of this thesis. I would also like to express my appreciation to Dr. L.Z. Florence for suggesting this project and his input and role as a committee member. I am also grateful to Dr. R.B. Hodgetts, and Dr. D. Gerson, who took the time to wade through my thesis and offer constructive criticism. I am indebted to the Genetics Department for extending full priviledges of departmental equipment and especially Dr. F.E. Nargang for his generosity in treating me as one of his own graduate students with respect to both lab space and technical advice. Lastly, I would like to thank both Dr. K. Morgan for his input in my analysis and interpretation of results and Mary Holmes who helped me tie it 'all together' in the form of a thesis.

Support for this project funded by the Canadian Forestry Service, the Natural Science and Engineering Research Council of Canada, and the Faculty of Graduate Studies and Research is gratefully acknowledged.

Table of Contents

Chapter — Page

I. INTRODUCTION 1

A. Statement of Problem 2

B. Actin Functions 4

THE ROLE OF ACTIN 4

Mammalian actins 4

Plant Actins 6

QUANTITATIVE AND QUALITATIVE ASPECTS OF ACTIN GENES 7

Number and locations of actin genes 7

Nucleotide and amino acid comparisons 7

Regions of high variability within the gene 8

Possible control sites 9

Introns and Exons 10

II. MATERIALS AND METHODS 12

A. DNA Isolation 12

Extraction and Isolation of Lodgepole Pine DNA 12

Recipes 15

B. Blot Hybridization 15

C. Partial Library Construction 16

D. DNA Sequencing 17

E. Sequence Analysis 18

III. RESULTS AND DISCUSSION 19

A. Southern Blot Analysis 19

B. Sequencing the PAcl-A Fragment 38

C. Comparisons of Actin Genes of Different Species 38

D. Localized Variation 48

E. Codon Positional Changes 52
F. Codon Useage 53
G. Intron 55
IV. CONCLUSION 58
V. LITERATURE CITED 61

LIST OF TABLES

TABLE		PAGE
1.	Nucleotide and Amino Acid Comparisons	44
2.	Percent of Codon Positional Changes	45
3.	Comparisons of Percent Amino Acid Homology with Weighted Amino Acid Substitutions	47
4.	Species Diversity Comparisons of Nucleotides and Amino Acids of the Carboxy Portion of Actin Proteins	49
5.	Constant vs. Variable Region of the Actin Gene	51
6.	Codon Usage in the Carboxy-Terminus of a Lodgepole Pine Actin Gene	54

LIST OF FIGURES

FIGURE		PAGE
1.	Autoradiograph of digested pine genomic DNA hybridized with ^{32}P-labelled SAc3.	21
2.	Sample of transformed populations screened with ^{32}P-labelled SAc3.	23
3.	Autoradiograph of undigested mini-plasmid preparations of populations Southern transferred and hybridized to nick-translated SAc3.	25
4.	Endonuclease restriction analysis of the plasmid, PAc1.	28
5.	Restriction map of PAc1.	30
6.	Autoradiograph of digested Pac1 DNA hybridized with ^{32}P-labelled SAc3.	32
7.	Autoradiograph of digested pine genomic DNA hybridized with ^{32}P-labelled PAc1.	34
8.	Autoradiograph of digested pine genomic DNA hybridized with ^{32}P-labelled pBR322.	37
9.	Physical map and sequencing strategy of pPAc1-A.	40
10.	Nucleotide sequence of the PAc1-A lodgepole pine actin fragment.	42
11.	Comparison of homology in the third intron of the plant actins.	56

LIST OF ABBREVIATIONS

Ala - Alanine

α - alpha

A.A. - amino acid

Asp - aspartine

Arg - arginine

Asn - Asparagine

β - beta

b.p. - base pair

CAAT box - consensus sequence of the form GGT/CCAATCT

Cap Site - consensus sequence of the form CCATACA

CsCl - cesium chloride

Cys - cysteine

dCTP - deoxycysteinetriphosphate

dGTP - deoxyquaninetriphosphate

dITP - deoxyinosinetriphosophate

DNA - deoxyribonucleic acid

EDTA - Ethylenediaminetetraacetic Acid (Disodium Salt)

EtBr - ethidium bromide

γ - gamma

Glu - glutamate

Gln - glutamine

Gly - glycine H-NET - high salt, EDTA, Tris Buffer (100mM Tris-Cl, pH 7.2, 150 mM NaCl, 1 mM EDTA)

His - histidine

Ile - isoleucine

Kb - kilobase

SDS - lauryl sulfate

Leu - leucine

Lys - lysine

mRNA - messenger ribonucleic acid

Met - methionine

ug - microgram

mg - milligram

ml - millileter

'-' - negative

N.T. - nucleotide

% - percent

Phe - phenylalanine

^{32}P - phosphorus 32

pg - picogram

'+' - positive

Pro - proline

RNase - ribonuclease

RNA - ribonucleic acid

Ser - serine

NaOAc - sodium acetate

NaCl - sodium chloride

TATA box - Hogness box or consensus sequence of the form TATAA/TAA/T

Thr - threonine

TE - Tris, EDTA buffer (10 mM Tris-Cl,pH 8.0, 1mM EDTA)

Trp - tryptophan

Tyr - tyrosine

UT - untranslated

Val - valine

I. INTRODUCTION

Traditional tree improvement programs generally entail complex breeding and orchard designs coupled with intensive management. These programs can be very costly in time and money with no guarantee that the hoped-for high genetic gains will be realized. The breeding strategies typically employed take long periods of time based on the age of reproductive maturity, the age when selections are made, and the number of generations required to achieve the breeding objectives. These breeding programs may also be limited by reproductive barriers and inbreeding problems during attempts to produce new gene combinations or transfer specific traits from one species to another (Faulkner, 1975). Attempts have been made to use techniques such as polyploidy and protoplast fusion to overcome reproductive incompatabilities with limited success (Shepard *et al.*, 1983). Strategies using inbreeding can result in lethality, inbreeding depression, narrowing of the genetic base, and the creation of monocultures susceptible to major diseases and/or insect epidemics.

Genetic manipulation of plants using molecular and cell culturing methods potentially could overcome some of these limitations of traditional tree improvement problems in the following ways: 1)insertion of small DNA fragments carrying one or a few genes avoids disruption of favourable gene combinations; 2)insertion of these fragments maintains and creates genetic diversity, and 3)recombinant DNA technology is not dependent on sexual reproduction. Thus, 'genetic engineering' offers the possibility of a dramatic solution to some of the inherent problems associated with forest tree improvement.

However, problems are also associated with plant molecular biology, because little is known about the identity and number of genes that are responsible for the expression of traits most economically important traits in plants. In addition, control of their expression is unknown. Therefore, the isolation and characterization of genes and understanding the underlying control mechanisms in higher plants are necessary before progress can be made employing these genetic techniques. Other problems associated

with genetic engineering involve transforming isolated DNA into plant tissues and cells and obtaining expression of desirable genes in foreign systems. Much progress has recently been made using *Agrobacterium tumefaciens* (e.g. Watson *et al.*, 1975) containing the Ti plasmids (e.g. Bevan and Chilton, 1982; Barton and Brill, 1983; Chilton, 1983). For instance, transformation of the phaseolin gene, coding for a major storage protein in bean, has been successfully transformed and expressed in tobacco plants (Netzer, 1984). Infection using *Agrobacterium* strains has been thought to be limited primarily to dicot crop plants. However, recent work by Olsen (1985) and Stomp *et al.* (1985) have shown that some of the *Agrobacterium tumefaciens* strains can infect and transform loblolly pine (*Pinus taeda* L.) germinants (reported by Olsen, W. and Stomp, A. at Molecular Genetics Working Party Conference October 1985, Avon Lake, Ohio).

A. Statement of Problem

This research was designed to isolate and characterize the actin gene(s) in lodgepole pine (*Pinus contorta* var. *latifolia* Engelm.), a widely distributed and commercial tree species in both Alberta and much of western North America. Actin was chosen because it is a highly conserved, ubiquitous protein (Williamson, 1980; Firtel, 1981; Shah *et al.*, 1982) that has been cloned and fully sequenced in many organisms. It has been suggested that most nucleotide alterations observed in actin that exist within and between species probably result in neutral mutations (e.g. Fyrberg *et al.*, 1981). Hence, it was expected that any organism's actin DNA could theoretically be used as a probe to locate the lodgepole pine actin gene(s), using DNA-DNA hybridization. The gene(s) could then be further defined by nucleotide sequence analysis. In addition, the number of genes per genome could be inferred by use of hybridization procedures. In this study, a portion of a lodgepole pine actin gene was located and isolated using the recombinant plasmid pSAc3 (obtained from R.B. Meagher, University of Georgia,

U.S.A.) as a probe. The SAc3 *HindIII* fragment contains an entire soybean actin gene with 5' and 3' flanking sequences and has been fully sequenced, and analyzed (Shah *et al.*, 1982). Therefore, pSAc3 (obtained from R.B. Meagher) was a convenient plasmid to use as a probe for lodgepole pine actin-related DNA.

Actin has been analyzed and characterized for species such as *Drosophila* (Fyrberg *et al.*, 1981), yeast (Gallwitz and Sures, 1980), *Dictyostelium* (Kindle and Firtel, 1978), humans (Ponte *et al.*, 1984; Engel *et al.*, 1981; Nakajima-Iijima *et al.*, 1985), rat (Zakut *et al.*, 1982) and Characean algae (Kershey and Wessels, 1976). Analyses of evolutionary relationships among nucleotide sequences, their encoded amino acids, and possible control mechanisms should be possible among all of these organisms and the lodgepole pine gene(s) to be analyzed in this study. Understanding underlying control mechanisms of plant genes is essential for gene expression in transformed cells. To date inferred TATA and Cap sequences have been located 'upstream' from the start codon of the actin gene in species such as yeast (Gallwitz and Sures, 1980), soybean (Shah *et al.*, 1982), *Oxytricha* (Kaine and Spear, 1982), and rat (Zakut *et al.*, 1982). The TATA sequence is generally located approximately 25 nucleotides before the starting point of transcription and may be involved in accurate positioning of the RNA polymerase II for initiation of transcription, whereas the Cap sequences are involved with capping mRNA that can subsequently be recognized by ribosomes to initiate translation. CCAAT sequences that have been implicated in the control of transcription have been found in actin genes of organisms such as *Oxytricha fallax* (Kaine and Spear, 1982), humans (Nakajima *et al.*, 1985), and rat (Nudel *et al.*, 1983). In addition, AATAAA polyadenylation sites have been located 'downstream' from the termination codon in other organisms such as sea urchin (Shuler *et al.*, 1983), human (Ponte *et al.*, 1984), and chicken (Bergsma *et al.*, 1985).

B. Actin Functions

THE ROLE OF ACTIN

Two forms of the actin gene, coding for seven different isoelectric proteins, have been found in mammals and birds (Vandekerckhove, J. and Weber, K., 1978a and 1978b; Bergsma *et al.*, 1985). Four of these proteins are present as alpha-actins (α-actins) and are found in the muscle tissues; vascular smooth, enteric smooth, cardiac striated, and skeletal striated of mammalian organisms. The muscle actins are easily diffentiated as they have a different N-terminal sequence from that of cytoplasmic actins (Fyrberg *et al.*, 1981).

The other three isoforms are cytoplasmic actins. The beta and gamma actins have been found in organisms such as *Dictyostelium* (Vandekerckhove and Weber, 1980), yeast (Gallwitz and Seidel, 1980; Gallwitz and Sures, 1980; Ng and Abelson, 1980), sea urchin (Shuler *et al*, 1983), soybean (Nagao *et al.*, 1981; Shah *et al.*, 1983), maize (Shah *et al.*, 1983), rats (Nudel *et al.*, 1983), and humans (Nakajima-Iijima *et al.*, 1985).

Mammalian actins

In mammalian organisms, α-actin is the prominent protein component of the thin muscle filaments. It is usually identified by its ability to react and bind to heavy mero-myosin (Pollard and Weihing, 1974; Weihing, 1976; Korn, 1976). The importance of cytoplasmic actin and its currently accepted roles in animal cells fall into four main categories, based on specific interactions. The first of these interactions occurs between actin filaments, producing a complex network, the 'cytoskeleton' (Brown *et al.*, 1976), which is thought to maintain cell shape. It should be noted that not all cytoplasmic actin is filamentous (Tilney, 1976; Tilney and Detmers, 1975). Actin can be found as a monomer (G-actin) which can interact with itself or ATP or it can be found in linear aggregates referred to as F-actin (Wegner, 1985). The second role is based on the

ability of actin to bind to some membrane proteins. It is thought that these interactions can affect protein mobility along the plasma membrane. For example, actin appears to reduce the mobility of spectrin (Fowler and Bennett, 1978; Goodman and Branton, 1978) and facilitates the mobility of other membrane proteins, resulting in localization of certain proteins into particular regions within the cell (Koch and Smith, 1978; Condeelis, 1974). In fact, a 110 kilodalton cytoskeletal protein has been identified that connects actin filaments to the plasma membrane in microvilli of intestinal epithelial cells (Glenney and Glenney, 1984). A third possible role, organelle immobilization (e.g. Seifriz, 1942), may result in production of a localized rigid gel formed by cross-reactions of actin filaments with non-myosin proteins (eg., Hartwig and Stossel, 1975). However, this process is only hypothetical as it is not understood and may not even occur *in vivo* (Williamson, 1980). The fourth role of actin is deduced from effects of an actin antibody and actin modulator protein on the Salamander *Pleurodeles waltlii*. When these substances were microinjected into living oocytes, transcription of lampbrush chromosomes was blocked, suggesting a close relationship between the state of nuclear actin and transcriptional activity (Scheer *et al.*, 1984).

Other studies of actin in organisms such as sea urchin (Garcia *et al.*, 1984), mouse (Waters *et al.*, 1985), and *Drosophila* (Fyrberg *et al.*, 1983) have analyzed actin patterns related to different tissues and cell types to various developmental stages. In general, different actins appear to have unique patterns of expression dependent on the tissue type and developmental stage analyzed. For example, *Drosophila* actins have three main patterns of expression correlated with muscle differentiation and reorganization. It is thought that actins responsible for these patterns are under different regulatory systems. Similar results were obtained from the sea urchin data. The three different patterns are labeled as muscle specific, embryo specific, and constitutive.

Plant Actins

Fewer data are available for the role of actin in plant cells. Systems similar to those in animals possibly exist in plants where actin is involved in many cellular activities such as cytoplasmic streaming, secretory processes, chloroplast movement, cytoskeletal maintenance, cell division, and possibly chromosome segregation (Korn, 1978). Actin may also be involved in interconnecting the nucleus with the cell membrane and distributing membrane proteins (e.g. Korn, 1978). Kersey and Wessells (1976) studied cytoplasmic streaming in characean algae and proposed that subcortical filaments containing actin and myosin generate force to move organelles within the cell either directly or indirectly. The filament-organelle associations were visually observed with electron microscopy where an average of 4-6 fibrils (or subcortical actin filament bundles) bridging lines of chloroplast subcortical fibrils were observed. These actin fibrils also appear to play a structural role as they link chloroplasts together. Similar associations between mitochondria and fibrils were observed, where the mitochondria may provide the required energy for cytoplasmic streaming. F-actin, which is a linear aggregate of the actin monomer, G-actin, has also been observed in conifer roots (Pesacreta *et al.*, 1982). Strong correlations have been made between the presence of microfilament actin bundles and cytoplasmic streaming (Pesacreta *et al.*, 1982). Two other possible functions of actin, whereby cell processes may be controlled, have been forwarded by Forer and Jackson (1975), and Gawadi (1971), who proposed that actin in spindle fibers affects chromosomal movement in mitosis. This was deduced by the occurrence of actin observed in spindles of both plant and animal mitotic cells. The final role actin may play in plant cells is related to the membrane protein-actin interactions already observed in animal cells. Membrane protein-actin interactions have not yet been directly observed in higher plants, and the role that actins may play in secretory processes remains ambiguous.

QUANTITATIVE AND QUALITATIVE ASPECTS OF ACTIN GENES

Number and locations of actin genes

The number and organization of actin genes varies among different species and appear to be unrelated to the degree of complexity among species (Firtel, 1981). To date, the yeast genome is the only organism carrying a single cytoplasmic actin gene, whereas actins from all other species analyzed belong to multigene families. *Caenorhabditis elegans* contains four actins, three of which are tightly clustered (Files *et al.*, 1983), human and rat genomes carry 20-25 actins, and *Dictyostelium* contains 15-20 gene copies of one cytoplasmic actin isoform. *Dictyostelium* actin genes occur in clusters with at least two linked actin genes separated by low copy spacer DNA of variable length (McKeown *et al.*, 1978). In contrast, the actin gene family in *Drosophila* consists of six genes coding for all three cytoplasmic protein isoforms, which are widely distributed throughout the genome as demonstrated by *in situ* hybridization mapping (Tobin *et al.*, 1980; Fyrberg *et al.*, 1981). Interestingly, *Drosophila* has no actin genes coding for the muscle alpha-actin proteins. Evidence from genomic hybridization data suggests that soybean actin is encoded in a small, multigene family having a minimum of ten cytoplasmic actin genes (Nagao *et al.*, 1981).

Nucleotide and amino acid comparisons

Varying degrees of conservation exist between different actins. For example, human, rat, bovine and chicken actins share 99.8% to 100% amino acid homology (Nakajima-Iijima *et al.*, 1985), whereas yeast shares 77% to 79% amino acid homology with soybean and maize (Shah *et al.*, 1983), respectively.

In general, the amino acid sequences of different actins within and between species show higher conservation than nucleotide comparisons. For instance, *Drosophila* and yeast actins share 86.4% amino acid homology but only 73% nucleotide homology

(Shah *et al.*, 1982). The two soybean actin proteins, SAc3 and SAc1, are 87.2% conserved when the amino acid sequences predicted from the DNA sequence are compared, whereas the nucleotide sequences have 77.2% homology. This trend can be explained by the prevalence of third position codon changes that are mostly silent with respect to amino acid replacements.

Regions of high variability within the gene

Two localized regions within the actin gene appear to be more diverse relative to the remainder of the coding sequence when different actins are compared.

The N-terminal region of different actin proteins among species is highly variable with the exceptions of residues 7-9 (Ala-Leu-Val) and residues 11-15 (Asp-Asn-Gly-Ser-Gly) which are homologous in yeast, *Dictyostelium*, *Drosophila*, and muscle and cytoplasmic specific actins. Soybean actin differs slightly from these sequences at 7-9 (Pro-Leu-Val) and 11-15 (Asp-Asn-Gly-Thr-Gly), where the altered amino acids are underlined (Pro from Ala and Thr from Ser respectively). These data suggest that residues 7-9 and 11-15 are structurally important. For all of these species, the N-terminal 3-4 residues are invariably acidic. With the exception of the residues 7-9 and 11-15 outlined above, the N-terminal nucleotides are highly variable within organisms and between species.

A second region of localized variation has been reported between codons 259 and 298 by Vandekerckhove and Weber (1978b, 1980). This regional variation is also evident when yeast and the soybean SAc3 are compared, except that variability occurs from residues 228 to 277.

In general, 3' and 5' flanking sequences of different actins are widely diverged between and within species. However, the beta actins of human and rat share more than 80% homology in both the 5' and 3' untranslated (UT) regions. Furthermore, the 3' UT sequence of human beta actin used as a radioactively-labeled probe hybridizes

strongly to chick beta actin mRNA (Ponte *et al.*, 1984). Ponte *et al.* (1984) suggested that this homology may be maintained by strong evolutionary pressure related to the possible regulatory role the 3'-UT region may play, although no specific role was suggested.

Possible control sites

Similarities in regulatory features have been observed in the 5' flanking sequences between many different actin genes. Sites with striking homology with the putative promoter TATA sequence or Hogness box (Gannon, 1979) have been found in the 5'-UT region. For instance, the yeast TATA sequence is located 54 base pairs "upstream" from the AUG start codon, whereas, the soybean Hogness box is located 98 base pairs "upstream" from the start codon. This TATA sequence is similar to that seen in other eukaryotic genes (Shah *et al.*, 1982) and may function as a promoter recognition site. In addition, the potential Cap sequence (C-C-A-T-A-C-A), as defined by Hentschel *et al.* (1980), appear 27 nucleotides "downstream" from the start of the T-A-T-A sequence in soybean and a potential CAAT box 61 nucleotides upstream from TATA occurs in the human beta actin gene (Nakajima-Iijima *et al.*, 1985).

Many of the 3'-UT regions of different actins share the sequence AATAAA found about 20 residues upstream from the 3'-terminal poly(A) track in actin mRNAs. This site is also found in other mRNA species such as alpha and beta globulin mRNA of rabbits and human, ovalbumin mRNA of chicken and the immunoglobulin light chain mRNA of mouse (Proudfoot and Brownlee, 1976). The AATAAA hexamer in the mRNA is important for cleavage and polyadenylation of mRNA (Birnstiel *et al.*, 1985).

Introns and Exons

Introns have been defined as transcribed DNA segments that are processed or excised by splicing together the flanking sequences or exons on either side (Lewin, 1985). Although actin gene exons are highly conserved, the number, locations, and nucleotide sequences of the introns vary. No introns have been found in *Dictyostelium* (McKeown *et al.*, 1978; Kindel and Firtel, 1978), nor one of the chicken beta actins (Bergsma *et al.*, 1985), nor *Oxytricha fallax* (Kaine and Spear, 1982), whereas five introns are present in the rat skeletal muscle actin (Zakut *et al.*, 1982) and seven in the human aortic smooth muscle actin (Engel *et al.*, 1981). It is interesting to note that the human beta actin, one of the chick beta actins, and the rat skeletal muscle actin all contain a single intron in the 5'-UT region when the DNA and mRNA sequences are compared.

Three introns of varying lengths occur at identical locations along the plant actin genes MAc1, SAc1 and SAc3 but share no nucleotide similarity with each other, either between introns of the same or different genes. One of the introns is found between codons 20 and 21, another between codons 355 and 356, and the other intron splits the glycine codon at position 150 (Shah *et al.*, 1982; Shah *et al.*, 1983). Yeast contains one intron between codons three and four whereas the chicken beta actin does not contain an intron within its coding sequence. *Drosophila* actin genes have one intron, but the position of each intron varies among the different actin genes (Fyrberg *et al.*, 1981). Zakut *et al.* (1982) suggested that intron site conservation may be used as an indicator of evolutionary relatedness. Similarily, Shah *et al.* (1982), proposed that positional intron conservation in the maize and soybean actins support the view that the plant actin genes may have evolved from a common ancestral actin gene containing a minimum of three introns.

Possible functions of introns have been investigated in yeast (Ng *et al.*, 1985), an organism that is very responsive in its growth to altered levels of actin expression, whereby the actin gene with its intron deleted was transformed into yeast cells containing

a mutant actin gene. No differences between the wild-type yeast cells and cells containing the actin gene with a deleted intron, were detected leading to the assumption that the intron plays no important role in the function of the yeast actin gene.

Gilbert (1978) proposed that the primary function of introns is to facilitate gene evolution, i.e., the more primitive a gene, the more introns it is likely to have. Therefore, many introns may have been deleted from *Drosophila* during evolution because there has been a long time period to excise introns from actin. It has been suggested (Zafar and Sodja, 1983) that actin is an ancient protein molecule originating in the prokaryotes. This is based on the observation that higher eukaryotes have higher numbers of actin isoforms when compared to prokaryotes.

II. MATERIALS AND METHODS

A. DNA Isolation

At the onset of this project, much effort was made at obtaining consistently high yields of high molecular weight DNA from lodgepole pine seedlings. Previous procedures reported for the isolation of pine DNA yielded up to 100 micrograms of DNA per gram fresh weight of tissue (e.g., Zimmet *et al.*, 1981; Miksche and Dhillon, 1981). Eventually, a protocol was developed where higher DNA yields were obtained, using solutions 1 and 2 (see recipes) as outlined in the procedure of Coen *et al.* (1982) for DNA isolation from *Drosophila*. This protocol, the outline of which follows, was reasonably rapid and yielded from 200 to 500 micrograms of DNA per gram fresh weight of pine seedling tissues. The resultant high molecular weight DNA was easily digested with many restriction endonucleases. This protocol was followed for the first half of the project until it was found that steps 3 to 8a could be omitted. Thus, following grinding of the seedling tissue and addition of solutions A and B, 0.96 grams per ml of CsCl could be added directly to the tissue 'slurry'. Ethidium bromide (EtBr) was added to a final concentration from 300 to 500 ug/ml and the DNA was centrifuged at 45,000 rpm (Beckman L8-55 centrifuge, VTi50 rotor) for a minimum of 12 hours at 15°C. Steps 8b and 9 were then followed. This quicker method yielded from 100 to 300 ug of relatively clean, high-molecular-weight DNA, free of RNA contamination.

Extraction and Isolation of Lodgepole Pine DNA

(Solutions 1 and 2 from Coen *et al.*, 1982.)

1. Approximately 10 grams (fresh weight) of two week old germinants were surface sterilized with a 1% hydrogen peroxide solution, followed by a wash in 3% detergent (we use "lux" liquid detergent) solution and a final rinse with distilled

water. Seedling tissues was then ground to a fine powder in liquid nitrogen, using a precooled mortar and pestle.

2. The ground tissue was divided into two 30 ml Corex tubes and sufficient Buffer 1 was quickly added to both tubes to thoroughly cover the tissue (about five ml/tube). An equal volume of Buffer 2, preheated to 65°C, was then added to each tube. The ratio of total solution volume to tissue volume (as a 'slurry') was about 2.5:1, respectively (see Recipes, below).

3. The samples were incubated in a 65°C water bath for 45 to 60 minutes. The solution was noticeably viscous at this point.

4. The tubes were spun at 12,000 g max for 10 minutes at 20°C (SS34 fixed-angle rotor in a RC-5B Sorval centrifuge).

5. The supernatant was discarded. Equal volumes of Buffer 1 and Buffer 2 (about two ml of each) were added to the tubes and the pellet was gently loosened. Each pellet was then reground in a small, preheated (65°C) mortar and pestle to yield a fine homogeneous 'slurry' which was replaced in a clean Corex tube and further incubated for 10 to 15 minutes at 65°C.

6. Following regrinding and incubation, SDS was added to 1% of the total volume (50ul of a 10% stock solution) and 200ul of a self-digested pronase (Sigma type XIV) solution (two mg/ml in H-NET) was added to each tube. The samples were then incubated overnight at 37°C. (Note: The added SDS is in addition to the SDS present in Buffer 2).

7. Equal volumes of water-saturated phenol (BDH chemicals) were added to each tube, mixed thoroughly by gently inverting the sample for 10 to 20 minutes, and then centrifuged at 12,000 g max., 20°C, for 10 minutes. The resultant solution contained three discrete layers: (i) the lower phenol phase which also contained a pellet of macerated tissue and cellular debris, (ii) the interphase containing various precipitated cellular constituents, and (iii) the upper, highly viscous aqueous phase, which contained most of the DNA. This upper phase was gently

removed and placed in a clean 30 ml Corex tube. After transferring the viscous upper layer to a clean Corex tube, four ml of TE buffer were added to the material remaining in the tube. The samples were then mixed by inversion. Following centrifugation, the upper viscous 'DNA' layer was added to the first 'DNA' fraction already collected. This re-extraction was repeated as many times as desired resulting in successively higher total DNA yields per preparation. This extraction has been repeated up to five times where it was noted that the aqueous phase is still noticeably viscous aqueous phase. The final pooled aqueous phase containing DNA was then re-extracted with phenol at least twice more, or until the phenol phase became clear.

8. Residual phenol was removed from the aqueous phase by extracting at least once with an equal volume of chloroform:isoamyl alcohol (24:1). The aqueous DNA fraction was collected after centrifugation at 12,000 g max, for four minutes at 20°C.

   a. Upon addition of 1/10 volume of 3M NaOAc and layering three volumes of 95% ethanol over the viscous DNA solution, DNA was spooled out and subsequently resuspended in five ml of TE buffer.

   b. Although the spooled DNA could be cleaved with restriction endonucleases, more consistent results were obtained if the DNA was further purified by a cesium chloride-etidium bromide (CsCl-EtBr) ultracentrifugation step, using 0.98 gm of CsCl added to each ml of DNA solution and addition of EtBr to a final concentration from 100 to 400 ug/ml. Following centrifugation at 54,000 rpm (Beckman L8-55 centrifuge, VTi65 rotor) for a minimum of six hours at 15°C, the DNA band was collected. EtBr was removed from the solution by five or six extractions with isopropanol previously saturated with NaCl-20 mM Tris

solution.

9. To remove CsCl, the solution was dialyzed overnight in four litres of TE buffer with at least 1 buffer change, preferably 1-4 hours after beginning dialysis.

10. Following dialysis the DNA solution, containing small amounts of RNA, was often stored over a few drops of chloroform at 4°C. Alternatively, any remaining RNA was removed by adding 20ul of a heat-treated RNase A solution (two mg/ml in H-NET) for each ml of solution. This step was followed by phenol and chloroform/ isoamyl alcohol extractions to remove RNase and residual phenol, respectively. Residual chloroform /isoamyl alcohol was removed by ethanol precipitation or dialysis, as described above.

Recipes

1. Buffer 1: 10 mM Tris, 60mM NaCl, 5% sucrose, 10mM EDTA, pH 7.5.
2. Buffer 2: 1.25% SDS, 0.3M Tris, O.1M EDTA, 5% sucrose, pH 9.0. (Note: The buffer recipes used in this DNA isolation protocol were obtained from Coen *et al.*, 1982. Buffer 2 was modified by deleting diethyl pyrocarbonate.)
3. TE Buffer: 10mM Tris-Cl, pH 8.0, 1mM EDTA.
4. H-NET: 100mM Tris-Cl, pH 7.2, 150 mM NaCl, 1 mM EDTA.
5. 10X TBE Buffer: 1M Tris-Borate, pH 8.3, 20mM EDTA (disodium salt).

B. Blot Hybridization

Electrophoresis of genomic DNA (10-15 ug) was carried out on 0.8% agarose gels (Maniatis *et al.*, 1982) containing EtBr at a concentration of 0.5 ug/ml, transferred to Biodyne membrane (Pall Corp.) or Gene Screen Plus (New England Nuclear) and

hybridized with radioactive probe using the protocol outlined by Klessig and Berry (1983). The probe was a genomic DNA fragment encoding soybean actin (SAc3) isolated from the pSAc3 plasmid (a gift of R.B. Meagher) and labelled using a technique developed by Feinberg and Vogelstein (1983, 1984) that radiolabels denatured DNA to high specific activities with Klenow, the large fragment of DNA Polymerase 1 (Bethesda Research Laboratories) and random oligonucleotides (Pharmacia) as primers.

C. Partial Library Construction

The results of the hybridization analysis suggested approximately six HindIII fragments sized from 0.9 to 7.6 kilobases (kb) carried actin sequences. Therefore, HindIII-digested genomic DNA was fractionated on a 10-40% sucrose gradient (Maniatis *et al.*, 1982) and dialyzed against TE buffer. Pine HindIII restriction fragments sized from 0.5 to 8.0 kb were ligated into HindIII digested dephosphorylated pUC19 plasmid vector (Norrander *et al.*, 1983), transformed into the *Escherichia coli* bacterial strain JM83 (Yanisch-Perron *et al.*, 1985), and propagated such that 201 discrete transformed populations, each consisting of approximately 3000 transformants, were maintained. Fifty percent of the transformants within each population contained pine inserts. Thus, 1500 unique recombinant clones were present in each population. pUC19 DNA was dephosphorylated with calf intestinal phosphatase (Boehringer Mannheim). Populations were allowed to grow to saturation in three ml of L-broth overnight, screened separately by dotting aliquots of each population on biodyne membrane, and subsequently the filter was hybridized with radioactive SAc3 DNA, which had been purified from pBR322 DNA sequences. Populations of transformants, rather than individuals were initially screened rather than individual transformants to overcome problems presented by the large genome size of lodgepole pine (12.2 pg or 11.1 X 10^9 bp/haploid genome). This screening protocol enabled large numbers of transformed cells to be quickly and efficiently screened, because as many as 53 populations (containing a total of about 80,000

transformants) could be screened on a single membrane. Mini-plasmid DNA preparations were done on eight out of a total 201 populations found to hybridize with SAc3. Electrophoresis was carried out on a 0.8% agarose gel, which was subsequently Southern transferred. The single population found to hybridize with the SAc3 probe was further examined by filter-colony hybridization to isolate recombinant plasmids containing pine actin sequences. Southern transfer, filter colony transfers and subsequent hybridizations were carried out as per accompanying Biodyne membrane manufacturers' instructions. Restriction endonucleases and T4 DNA ligase were obtained from New England Biolabs and Betheseda Research Laboratories and were used in accordance with supplier's instructions. α-^{32}PdCTP was obtained from New England Nuclear. Plasmid DNA was isolated as described by Kahn *et al.*, 1979.

D. DNA Sequencing

DNA sequences were determined by the method of Sanger *et al.* (1977), using the M13 derivatives mp18 and mp19 (Norrander *et al.*, 1983) for the production of single stranded DNA templates. JM103 bacterial cells (Messing, 1983) were transformed using the procedure of Norgard *et al.*, 1978. Single-stranded DNA isolated from the transformed cells was primed with synthetic 17-mer obtained from New England Nuclear Laboratories. DNA sequencing reactions were carried out at 45°C and run on 6% pH 8.3 thin polyacrylamide gels (Sanger and Coulson, 1978). Compression of bands on the gel caused by secondary structure effects in the DNA template were overcome by replacing dGTP with dITP (deoxyinosine triphosphate) in the reaction mixtures. Secondary structure problems are less likely to occur because I:C base-pairing is weaker than G:C pairing (Sanger *et al.*, 1982). The computer programs of Staden (1977 to 1980b) were used to analyze the DNA sequence.

E. Sequence Analysis

Amino acid comparisons among the coding sequences of different actin genes was done using the Microgenie Analysis Program (Beckman Instruments). Intron homology comparisons were also done using the Microgenie Program, whereas nucleotide comparisons were done without the aid of a computer.

III. RESULTS AND DISCUSSION

A. Southern Blot Analysis

Detection of the pine actin gene family was achieved by hybridization of pine genomic DNA with a ^{32}P-labeled soybean actin gene, SAc3, as the probe. In single digests with the restriction enzymes *HindIII* and *EcoRI*, a minimum of approximately six and eight actin-related restriction fragments, respectively, were revealed (Figure 1). The results suggest that pine actin exists as a multigene family, although it is difficult to estimate the number of genes involved from these data alone.

Based on these results it was decided to clone the *HindIII* fragments into a bacterial plasmid for further analysis. Because most of the hybridizing bands of pine DNA were between 0.5 and 8.0 kb, *HindIII*-digested pine DNA sized to these limits was ligated into the bacterial plasmid pUC19. The ligated DNA's were transformed into bacterial cells, and 201 separate populations of transformants were maintained as described in Materials and Methods. Each of these populations was screened for the presence of pine actin sequences using the SAc3 probe. Figure 2 shows an example of 53 of these 201 screened populations. Populations #73 and #86 were chosen for further testing from this particular sample in addition to six other putative positive populations. Undigested plasmid DNA prepared by the "quick screen" method from these eight possible positives was electrophoresed, transferred to Biodyne membrane and hybridized with the SAc3 probe. Population #86 was found to hybridize strongly with the SAc3 probe (Figure 3), and was subsequently examined by filter colony hybridization of individual isolates from the population. This resulted in the isolation of a single clone with a pine actin containing *HindIII* restriction actin fragment. This plasmid was designated pPAc1. The other seven populations gave visually weaker hybridization responses when compared to population #86. It seems likely that these weaker responses resulted from the presence of low amounts of the 4.2 kb pBR322 DNA in the 3.0 kb SAc3 DNA isolated from the pSAc3 plasmid by extraction from a low

Figure 1

Autoradiograph of a southern transfer of pine genomic DNA digested with restriction enzymes and hybridized with nick-translated SAc3. Lane 1 - Pine DNA digested with *HindIII*. Lane 2 - Pine DNA digested with *EcoRI*. Lane 3 - Pine DNA digested with *HindIII* and *EcoRI*. Lane 4 - Pine DNA digested with *EcoRV*. Lane 5 - pSAc3 DNA digested with *HindIII*. The numbers next of the figure indicated the sizes of restriction fragments in kilobases. Residual pBR322 contamination present when the SAc3 fragment was isolated from pSAc3 can be detected in lane 5. The 4.2 kb linear pBR322 plasmid from the pSAc3 control hybridizes only slightly less intensely when compared to the 3.0 kb SAc3-Sac3 hybridization.

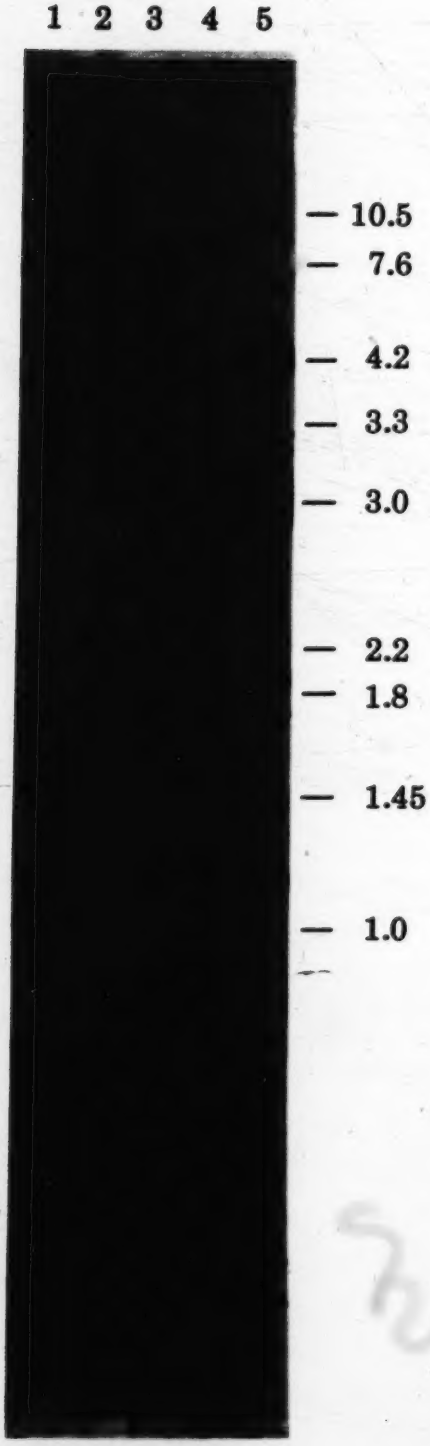

1 2 3 4 5
10.5
7.6
4.2
3.3
3.0
2.2
1.8
1.45
1.0

Figure 2

Sample of populations numbered 54 to 106 of transformants screened with nick-translated SAc3 DNA. Populations are numbered in increasing consecutive increments from those numbers indicated at the top of each column. Positive and negative controls were streaked in the shape of X's and are represented by '+' and '-', respectively. The '+' control consisted of JM83 cells containing pSAc3 plasmid, whereas the '-' control contained the pUC19 vector. Populations #73 and #86 were subsequently tested from this particular sample.

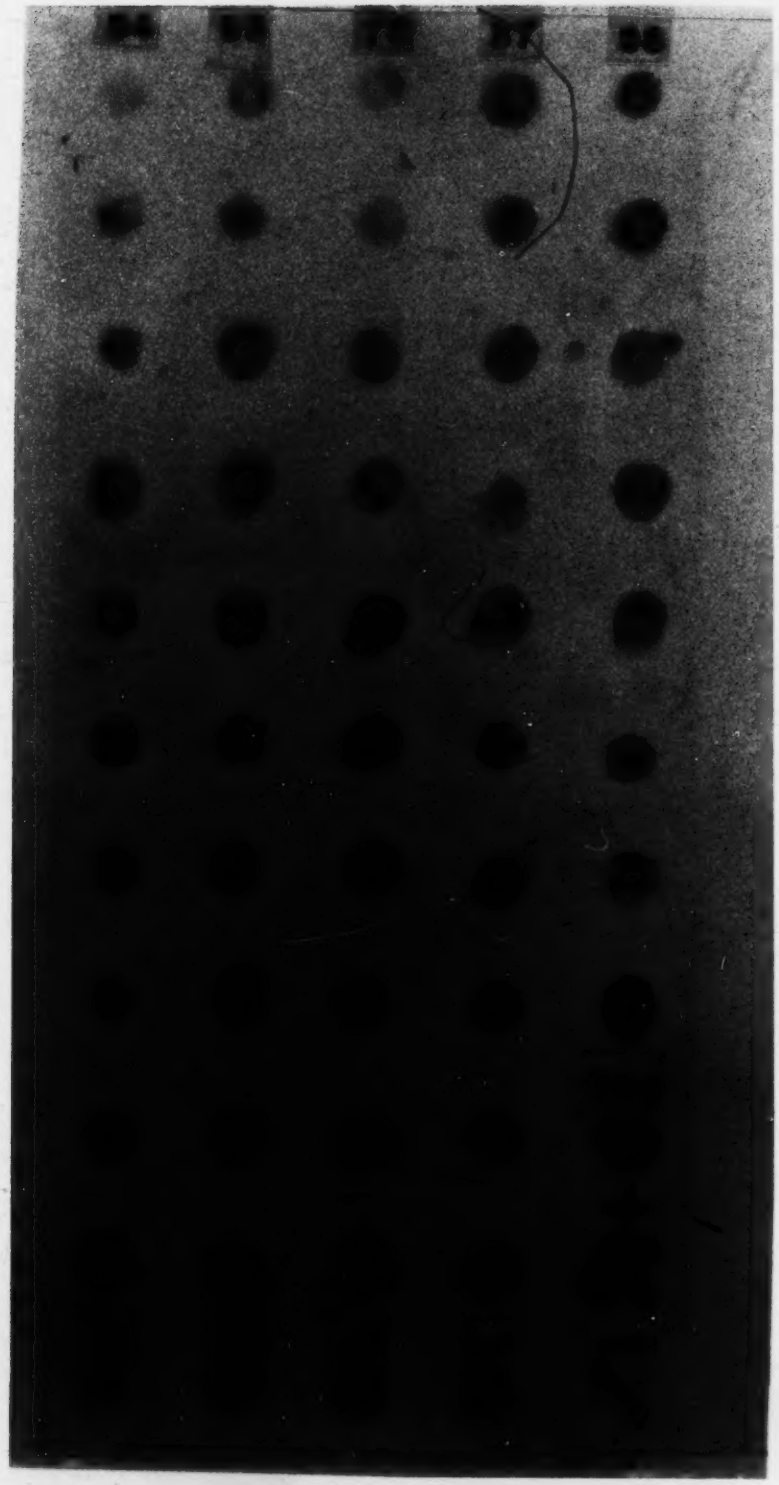


Figure 3

Autoradiograph of undigested mini-plasmid preparations of populations electrophoresed, Southern transferred and hybridized to nick-translated SAc3. Populations 195, 192, 182, 86, 73, 44, 11, and 8 are in lanes one to eight, respectively. Population #86 (lane 4) gave the most intense signal of all eight populations.

1 2 3 4 5 6 7 8

temperature melting agarose gel. The small size differences between the pBR322 and SAc3 DNA bands when subjected to electrophoresis resulted in the isolation of impure SAc3 DNA. This contaminating pBR322 DNA hybridized strongly to the pUC19 sequence in the recombinant plasmids, since pUC19 and pBR322 contain large regions of homology (Norrander *et al.*, 1983). In addition to the pSAc3 *HindIII* digestion in Figure 1, the presence of residual pBR322 contamination in SAc3 DNA isolates can be observed in the 'negative' control in Fig. 2. This negative control consists of JM83 cells containing the pUC19 plasmid, whereas the positive control consists of JM83 cells containing pPAc1. Although the 'negative' control response is not as intense as the 'positive' control, it still results in a stronger response than expected for DNA not containing actin when hybridized with SAc3. On this basis, the seven remaining populations were not further analyzed.

Restriction mapping experiments revealed that pPAc1 carried a pine DNA *HindIII* restriction fragment approximately 1.67 kb in size and resulted in the identification of a few endonuclease restriction sites contained on this fragment (designated PAc1) (Figure 4). A preliminary restriction map of PAc1 based on the digestion in Figure 4, is illustrated in Figure 5. The gel shown in Fig. 4 was subsequently transferred to Biodyne membrane and hybridized with SAc3 indicating that the pine actin sequences homologous to SAc3 were located on a 0.75kb *HindIII-EcoRI* subfragment of the original 1.67kb *HindIII* fragment (Figure 6). This *HindIII-EcoRI* fragment was subsequently recloned into pUC19 to give plasmid pPAc1-A. The residual contamination of the probe with pBR322, mentioned above, is evident in lanes one to six of Fig. 6, where hybridization to the 2.7 kb linear pUC19 DNA fragment occurs. Hybridization to the higher molecular weight bands likely resulted from partial digests.

The 1.67kb *HindIII* fragment described above was purified by electrophoresis of pPAc1 DNA digested with *HindIII* on a low temperature melting 0.8% agarose gel, and used to probe further blots of genomic pine DNA (Figure 7). As described later, PAc1 contains 483 nucleotides of the carboxy terminal portion of a pine actin gene. Since the

Figure 4

Endonuclease restriction analysis of the plasmid PAc1. PAc1 was digested with *HindIII*, excising the 1.67 kb lodgepole pine actin fragment from the 2.7 kb pUC19 vector in lanes 3 to 11. The plasmid band is present in all further double and triple digestions because it does not contain any restriction sites for the enzymes subsequently used in lanes 3 to 10. The PAc1 fragment was further restricted as follows. Lane 3 - *EcoRV* and *BglII*. Lane 4 - *EcoRI* and *BglII*. Lane 5 - *BglII*. Lane 6 - *EcoRV* and *EcoRI*. Lane 7 - *EcoRV*. Lane 8 - *EcoRI* and *BamHI*. Lane 9 - *BamHI*. Lane 10 - *EcoRI*. Lane 1 contains *HindIII* digested lambda DNA and lanes 2 and 12 contain *HindIII* and *EcoRI* digested lambda DNA. The *HindIII*, *HindIII* and *EcoRI* molecular weight standards are shown in kilobases along the right-hand side of the figure.

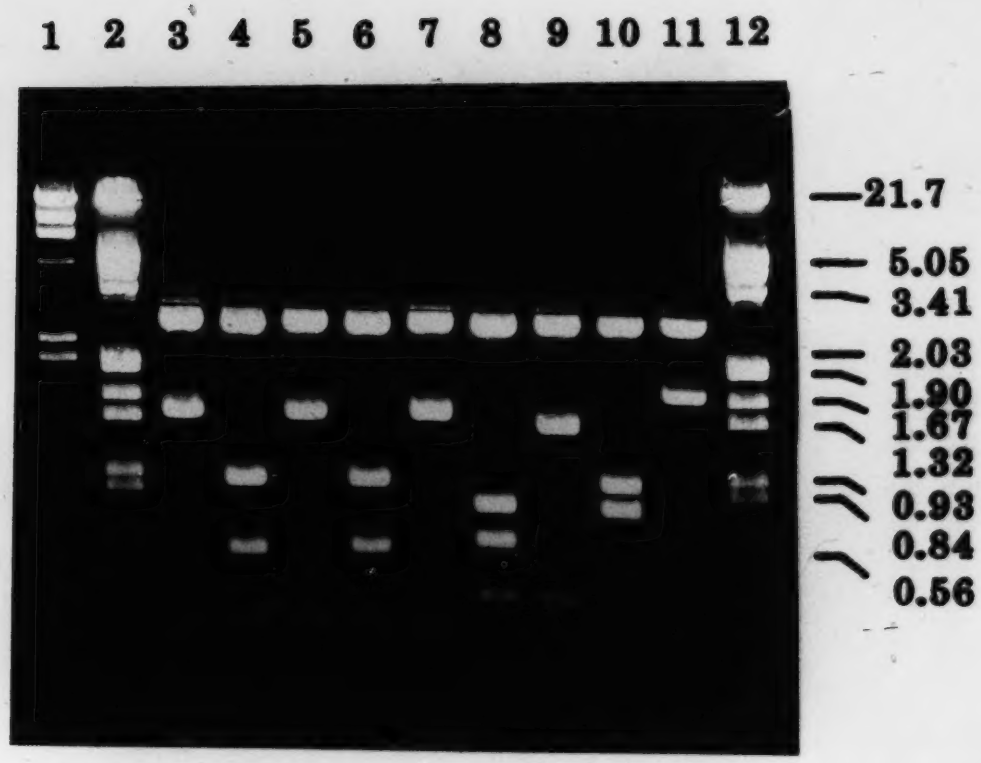

1 2 3 4 5 6 7 8 9 10 11 12
21.7
5.05
3.41
2.03
1.90
1.67
1.32
0.93
0.84
0.56

Figure 5

Preliminary restriction map of the lodgepole pine PAc1 actin fragment deduced from the restriction patterns in Figure 4. Approximate distances of the restrictions sites are indicated along the bottom of the figure in number of nucleotides.

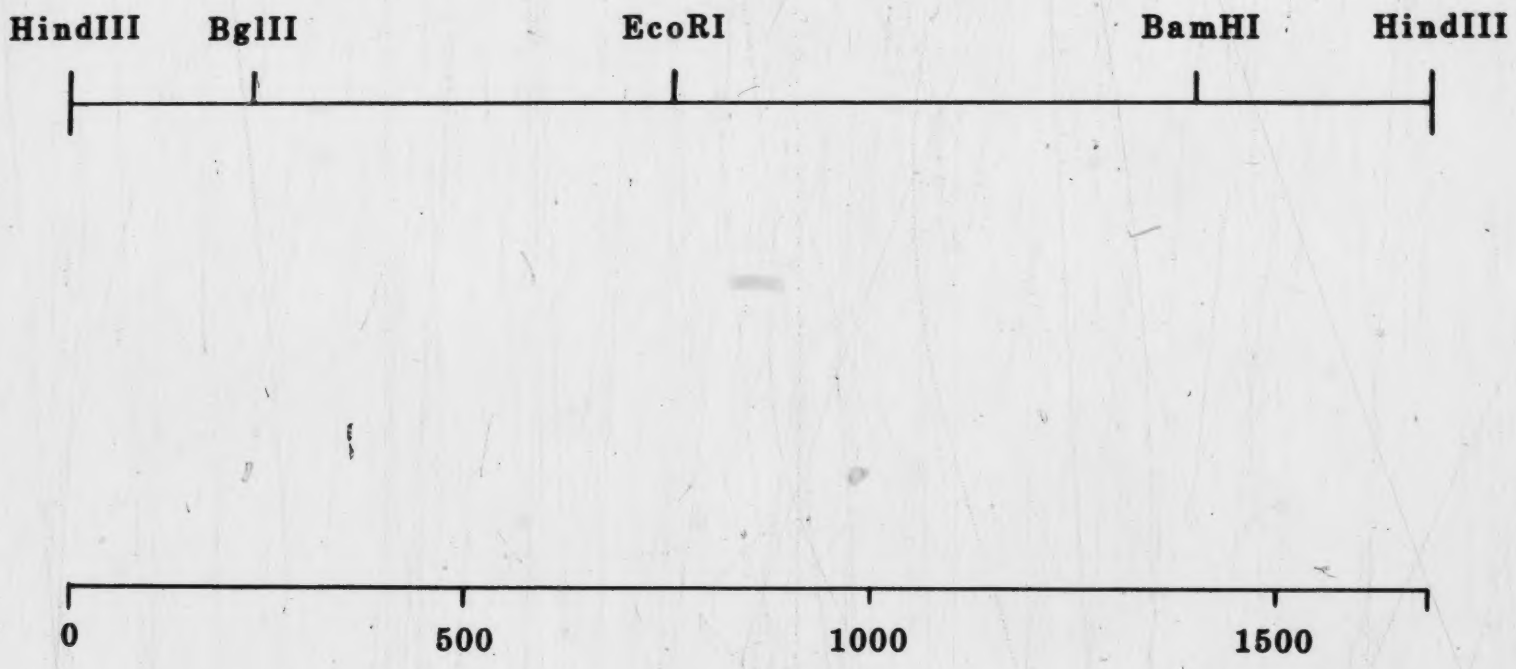

HindIII
BglII
EcoRI
BamHI
HindIII
0
500
1000
1500

Figure 6

PAc1 restriction fragments from the agarose gel shown in Figure 4, transferred to Biodyne membrane and hybridized to ^{32}P - labelled soybean probe, SAc3. The restriction enzymes used are as follows: Lane 1 - *EcoRV* and *BglII*. Lane 2 - *EcoRI* and *BglII*. Lane 3 - *BglII*. Lane 4 - *EcoRI* and *EcoRV*. Lane 5 - *EcoRV*. Lane 6 - *EcoRI* and *BamHI*. Lane 7 - *BamHI*. Lane 8 - *EcoRI*. Lane 9 - *HindIII*. The portion of PAc1 containing actin sequences was deduced from this autoradiogram. For example, the smaller *HindIII-EcoRI* fragment observed in lane 8(approximately 0.8 kb) contains actin coding sequences whereas the 0.9 kb fragment either carries very few or none at all. The response of the 230 bp fragments from *HindIII, EcoRI, BglII* (Lane 4)and *HindIII, EcoRI, EcoRV* (Lane 6) digests in addition to the larger 500 bp fragments indicate that *BglII* and *EcoRV* are contained within the actin coding sequence whereas *BamHI* likely is not. The occurrence of bands at molecular weights of 2.7 kb and higher in lanes 1 to 6 result from hybridization of residual contamination of pBR322 vector in the SAc3 probe isolated from pSAc3 on a low temperature melting point gel. Unequal hybridization of this residual contamination most likely result from an uneven transfer of DNA to the Biodyne membrane. The lambda *HindIII-EcoRI* molecular weight standards are shown in kilobases along the right-hand side of the figure.

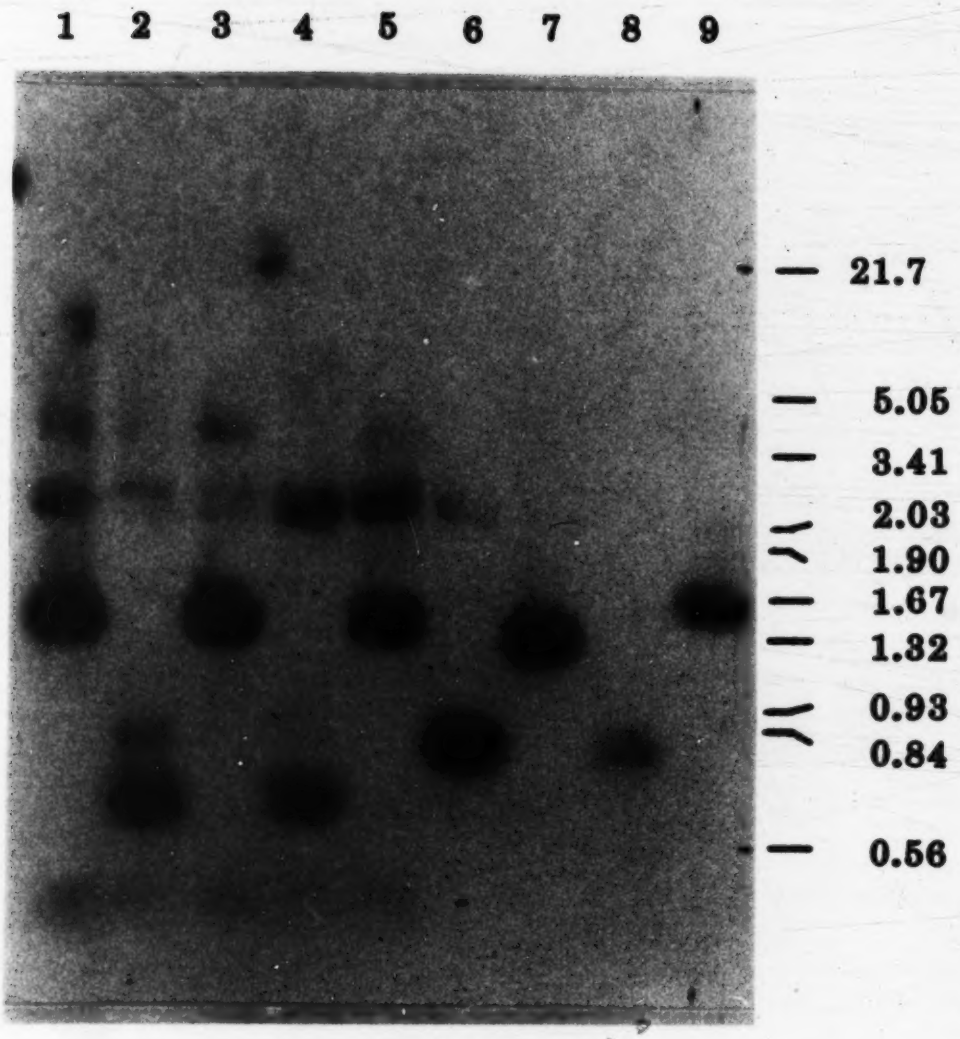
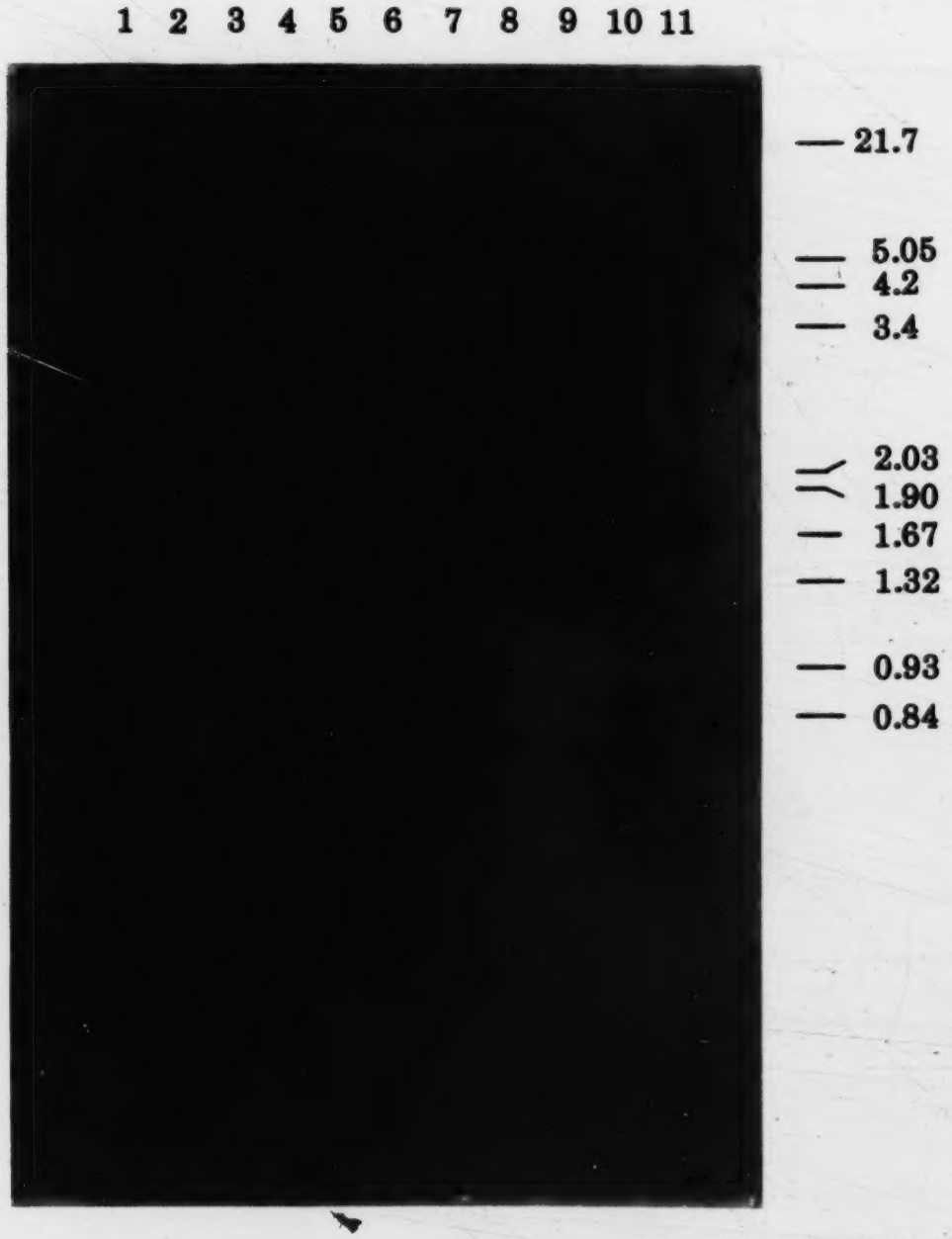

1 2 3 4 5 6 7 8 9
21.7
5.05
3.41
2.03
1.90
1.67
1.32
0.93
0.84
0.56

Figure 7

Autoradiograph of a Southern transfer of lodgepole pine genomic DNA digested with restriction enzymes and hybridized with ^{32}P-labelled PAc1 DNA. Lane 1 - pPAc1 digested with the restriction enzymes *HindIII*, *EcoRI*, and *BglI*. Lane 2 - pSAc3 DNA digested with *HindIII*. Lane 3 - Pine DNA digested with *HindIII* and *MspI*. Lane 4 - Pine DNA digested with *MspI*. Lane 5 - Pine DNA digested with *HindIII* and *HaeIII*. Lane 6 - Pine DNA digested with *HaeIII*. Lane 7 - Pine DNA digested with *HindIII* and *BglI*. Lane 8 - Pine DNA digested with *BglI*. Lane 9 - Pine DNA digested with *HindIII* and *EcoRI*. Lane 10 - Pine DNA digested with *EcoRI*. Lane 11 - Pine DNA digested with *HindIII*. The numbers along the right-hand side of the figure are the restriction fragment sizes in kilobases.


1 2 3 4 5 6 7 8 9 10 11
21.7
5.05
4.2
3.4
2.03
1.90
1.67
1.32
0.93
0.84

entire coding sequence of the SAc3 gene is 1128 nucleotides, PAc1 only contains about 43% of the pine actin coding region. Thus, fewer pine genomic restriction fragments are expected to contain nucleotide sequences homologous to the actin sequences contained on PAc1, relative to SAc3, and the number of these bands should closely correlate with the number of actin genes present in the pine genome. However, it is difficult to determine the absolute number of bands present on either of the blots hybridized with SAc3 or PAc1 because of the low resolution, particularily for the higher molecular weight fragments. In addition, it is possible that high molecular weight fragments carry more than one copy of actin. However, the resolution is adequate to differentiate a minimum of about nine *HindIII* bands and eight *EcoRI* bands when digested pine genomic DNA is hybridized with PAc1. It is possible that some of these *HindIII* and *EcoRI* bands potentially reflect partial digestion products. This would result in artificially high numbers of restriction fragments containing actin sequences. For all these reasons, it is difficult to calculate precisely the number of actin genes in pine. Assuming, complete digestion, the pine multigene family only can be estimated to contain a minimum of eight actin genes from the data in Figure 7.

In addition to hybridizing pine genomic DNA with the SAc3 and PAc1 probes, further hybridization using ^{32}P-labelled pBR322 was done to ascertain if the genomic DNA was contaminated with any plasmid vectors commonly used in our laboratory. To determine this, the same Southern transfer probed with PAc1 (autoradiograph shown in Figure 7) was washed and reprobed with the pBR322 probe (Figure 8). It is obvious from the autoradiograph presented that different pBR322 - related plasmids are present in the genomic DNA. Furthermore, the bands in Figure 8 correspond to some of the same bands observed in Figure 7. For example, the same 2.7 kb and 4.2 kb bands are present in both autoradiographs. Thus, it is apparent that contamination of pBR322 related plasmids account for some of the bands observed in Figure 7 that were initially assumed to be responses to hybridization of actin genes. It therefore becomes even more difficult to determine the number of genes contained in the lodgepole pine actin

Figure 8

Autoradiograph of the same Southern transfer hybridized with PAc1 (Figure 7) that was washed and reprobed with ^{32} P-labelled pBR322 DNA. Some of the bands, such as the 2.7 and 4.2 kb bands, correspond to bands observed when ^{32} P-labelled PAc1 DNA is used in the hybridization reaction. However, most of the other plasmid bands do not correspond to those observed in the autoradiograph depicted in Figure 7. The lodgepole pine genomic DNA and controls were digested as follows. Lane 1 - pPAc1 digested with *HindIII*, *EcoRI*, and *BglII*. Lane 2 - pSAc3 digested with *HindIII*. Lane 3 - Pine genomic DNA digested with *HindIII* and *MspI*. Lane 4 - Pine genomic DNA digested with *MspI*. Lane 5 - Pine genomic DNA digested with *HindIII* and *HaeIII*. Lane 6 - Pine genomic DNA digested with *HaeIII*. Lane 7 - Pine genomic DNA digested with *HindIII* and *BglII*. Lane 8 - Pine genomic DNA digested with *BglII*. Lane 9 - Pine genomic DNA digested with *HindIII* and *EcoRI*. Lane 10 - Pine genomic DNA digested with *EcoRI*. Lane 11 - Pine genomic DNA digested with *HindIII*. The numbers along the right-hand side of the figure are the restriction fragment sizes in kilobases.

1 2 3 4 5 6 7 8 9 10 11 12
23.1
9.5
6.4
5.0
4.4
3.5
1.98
1.90
1.59

family. Despite the difficulties of determining the absolute number of hybridizing bands, the data in Figures 1 and 7 are in many respects consistent with those in Figure 4. For example, Southerns blots containing *HindIII* digested pine genomic DNA and hybridized with either SAc3 or PAc1 contain a 1.67 kb band. This 1.67 kb fragment verifies the identity of PAc1-A as a clone derived from the pine genome. Similarily, pine genomic DNA digested with *HindIII* and *EcoRI* result in 0.75 and 0.95 kb fragments; *HindIII* and *BglII* digests result in 1.45 and 0.2 kb fragments; and *HindIII* and *MspI* result in 0.45 and 1.2 kb fragments when hybridized with PAc1 which correspond to the restriction sites contained in PAc1.

B. Sequencing the PAc1-A Fragment

Both strands of PAc1-A were fully sequenced as shown in Figure 9. The DNA sequence of the 750 bp HindIII-EcoRI fragment (PAc1-A) that hybridizes to SAc3 is presented in Figure 10. In positions where secondary structures in the DNA made the sequence unreadable, the appropiate clones were re-sequenced using dITP in the reaction mixture rather than dGTP (Sanger *et al.*, 1982). Comparison of the PAc1-A sequence to the published SAc3 sequence (Shah *et al.*, 1982) revealed that this pine DNA fragment contained coding sequences corresponding to amino acids 216 through to 376 of the soybean gene, SAc3. This represents the carboxy terminal portion of the gene. The coding sequence is interrupted by a single intron 130 nucleotides long that occurs between codons 355 and 356.

C. Comparisons of Actin Genes of Different Species

In order to evaluate the evolutionary relationship between the actin gene of pine and those of other species, the entire PAc1-A coding sequence was compared to the previously published coding sequences of soybean actins SAc3 and SAc1 (Shah *et al.*,

Figure 9

Physical map and sequencing strategy of the 0.75 kb *HindIII-EcoRI* fragment of pPAc1-A. The SAc3 amino acids (A.A.) to which the lodgepole pine actin fragment correspond, the locations of the intron, and the region of 3' flanking sequence are shown. The relative distance in basepairs have been drawn colinearly beneath the structural map. Sequencing was done through all the restriction sites used for subcloning into M13.

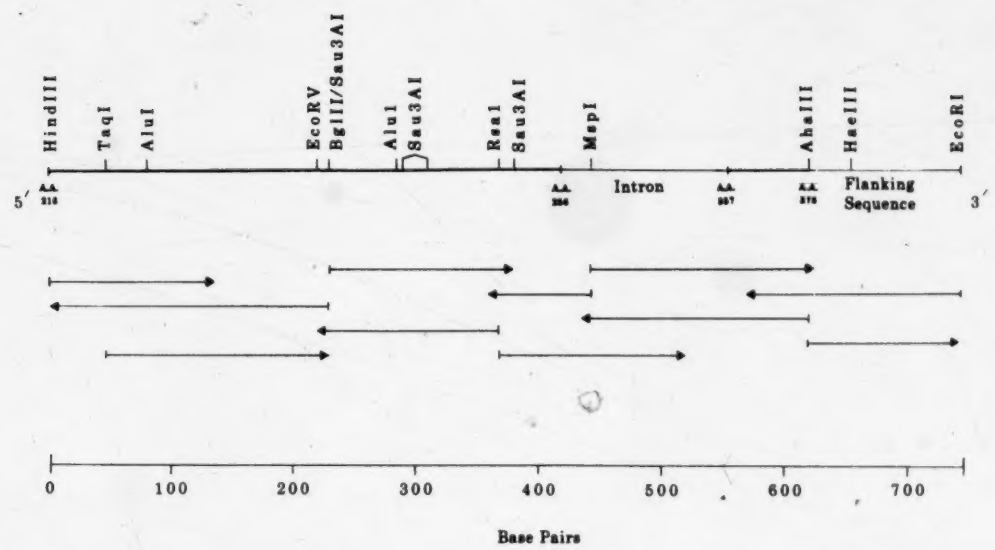
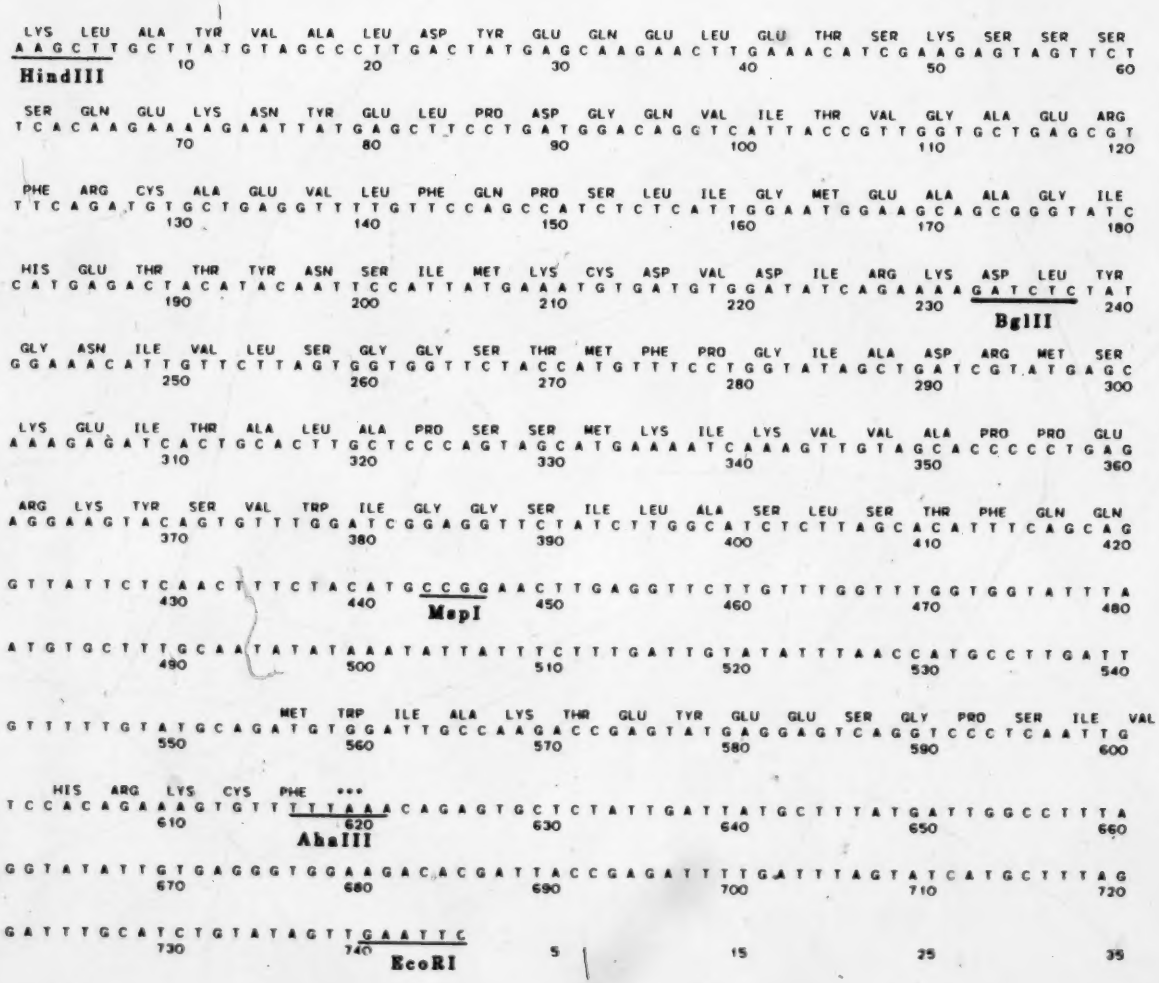

HindIII
TaqI
AluI
EcoRV
BglII/Sau3AI
AluI
Sau3AI
RsaI
Sau3AI
MspI
AhaIII
HaeIII
EcoRI
5′
3′
Intron
Flanking
Sequence
0
100
200
300
400
500
600
700
Base Pairs

Figure 10

Nucleotide sequence of the PAc1-A lodgepole pine actin fragment. This fragment contains the carboxy-terminal portion of a pine actin and includes one intron and 3' flanking sequences.


LYS LEU ALA TYR VAL ALA LEU ASP TYR GLU GLN GLU LEU GLU THR SER LYS SER SER SER
A A G C T T G C T T A T G T A G C C C T T G A C T A T G A G C A A G A A C T T G A A A A C A T C G A A G A G T A G T T C T
HindIII
10 20 30 40 50 60
SER GLN GLU LYS ASN TYR GLU LEU PRO ASP GLY GLN VAL ILE THR VAL GLY ALA GLU ARG
T C A C A A G A A A A G A A T T A T G A G C T T C C T G A T G G A C A G G T C A T T A C C G T T G G T G C T G A G C G T
70 80 90 100 110 120
PHE ARG CYS ALA GLU VAL LEU PHE GLN PRO SER LEU ILE GLY MET GLU ALA ALA GLY ILE
T T C A G A T G T G C T G A G G T T T T G T T C C A G C C A T C T C T C A T T G G A A T G G A A G C A G C G G G T A T C
130 140 150 160 170 180
HIS GLU THR THR TYR ASN SER ILE MET LYS CYS ASP VAL ASP ILE ARG LYS ASP LEU TYR
C A T G A G A C T A C A T A C A A T T C C A T T A T G A A A T G T G A T G T G G A T A T C A G A A A A G A T C T C T A T
190 200 210 220 230 240
BglII
GLY ASN ILE VAL LEU SER GLY GLY SER THR MET PHE PRO GLY ILE ALA ASP ARG MET SER
G G A A A C A T T G T T C T T A G T G G T G G T T C T A C C A T G T T T C C T G G T A T A G C T G A T C G T A T G A G C
250 260 270 280 290 300
LYS GLU ILE THR ALA LEU ALA PRO SER SER MET LYS ILE LYS VAL VAL ALA PRO PRO GLU
A A A G A G A T C A C T G C A C T T G C T C C C A G T A G C A T G A A A A T C A A A G T T G T A G C A C C C C C T G A G
310 320 330 340 350 360
ARG LYS TYR SER VAL TRP ILE GLY GLY SER ILE LEU ALA SER LEU SER THR PHE GLN GLN
A G G A A G T A C A G T G T T T G G A T C G G A G G T T C T A T C T T G G C A T C T C T T A G C A C A T T T C A G C A G
370 380 390 400 410 420
G T T A T T C T C A A C T T T C T A C A T G C C G G A A C T T G A G G T T C T T G T T T G G T T T G G T G G T A T T T A
430 440 450 460 470 480
MspI
A T G T G C T T T G C A A T A T A T A A A T A T T A T T T C T T T G A T T G T A T A T T T A A C C A T G C C T T G A T T
490 500 510 520 530 540
MET TRP ILE ALA LYS THR GLU TYR GLU GLU SER GLY PRO SER ILE VAL
G T T T T T G T A T G C A G A T G T G G A T T G C C A A G A C C G A G T A T G A G G A G T C A G G T C C C T C A A T T G
550 560 570 580 590 600
HIS ARG LYS CYS PHE ***
T C C A C A G A A A G T G T T T T T A A A C A G A G T G C T C T A T T G A T T A T G C T T T A T G A T T G G C C T T T A
610 620 630 640 650 660
AhaIII
G G T A T A T T G T G A G G G T G G A A G A C A C G A T T A C C G A G A T T T T G A T T T A G T A T C A T G C T T T A G
670 680 690 700 710 720
G A T T T G C A T C T G T A T A G T T G A A T T C
730 740 5 15 25 35
EcoRI

1982 and Shah *et al.*, 1983, respectively) maize actin MAc1 (Shah *et al.*, 1983), yeast (Gallwitz and Sures, 1980) and chicken beta actin (Kost *et al.*, 1983). The results of the comparisons are shown in Table 1. To date, the sequences of the soybean and maize actin genes are the only ones published for plant species. Yeast was chosen for the comparison because it is evolutionarily distant to soybean, maize and pine. The chicken actin sequence represented a different, unrelated species evolutionarily distant from both the higher plant and yeast species. Thus, members of the plant, animal, and fungal kingdoms are represented in this study. The amino acid differences given in the table were based on the total observed substitutions between any two genes. Cases where amino acids were unknown were treated as no changes. The nucleotide differences were based on total observed substitutions between any two genes only where nucleotides were known. For example, the SAc3 gene has three codons each with one unknown nucleotide. Thus, these amino acids were considered as matches in the amino acid comparisons and the unknown nucleotides were not used in the nucleotide comparisons for calculations of percent homology. Amino acid insertions and deletions were treated as one amino acid change regardless of how many amino acids were added or deleted at any one location. Nucleotide insertions and deletions were not considered in percent homology calculations.

In all the comparisons shown in Table 1, higher levels of homology were consistently observed at the amino acid level in contrast to the nucleotide level. This observation can be accounted for by the predominance of third-base codon substitutions, which are mostly silent (Table 2). Generally, the homology values in Table 1 correlate with evolutionary distances despite the high conservation of the actin protein. Yeast, which is a distant organism from the plant and chicken species, has the least amount of homology to pine. This contrasts with the pine-soybean and pine-maize comparisons, where higher homology values are observed among these more closely related species. Perhaps suprisingly, nucleotide and amino acid homology values of pine-chicken comparisons fall midway between those of pine-SAc1, pine-MAc1, and pine-yeast. It is

TABLE 1: Nucleotide and Amino Acid Comparisons

		NUCLEOTIDE		AMINO ACID	
		Total Changes	%Homology	Total Changes	%Homology
PAc1-A vs.	SAc1	109/481	77.3	24/161	85.1
	SAc3	110/483	77.2	15/161	90.7
	MAc1	120/483	75.2	25/161	84.5
	Chicken	144/483	70.2	29/161	82.0
	Yeast	167/483	65.4	35/161	78.3
SAc1 vs.	SAc3 (i)	253/1120	77.4	56/378	85.2
	(ii)	112/483	76.7	22/161	86.3
	MAc1 (i)	292/1119	73.9	65/378	82.8
	(ii)	123/479	74.4	32/161	80.1
	Chicken (i)	307/1124	72.7	71/378	81.2
	(ii)	140/481	70.9	35/161	78.3
	Yeast (i)	355/1117	68.2	89/378	76.5
	(ii)	153/475	68.3	40/161	75.2
SAc3 vs.	MAc1 (i)	253/1120	77.4	52/377	86.2
	(ii)	110/481	77.1	24/161	85.1
	Chicken (i)	286/1112	74.2	56/377	85.1
	(ii)	143/480	70.0	28/161	82.6
	Yeast (i)	332/1115	70.2	74/377	80.4
	(ii)	153/483	68.3	34/161	78.9
MAc1 vs.	Chicken (i)	277/1124	75.4	56/376	85.1
	(ii)	137/483	71.6	33/161	79.5
	Yeast (i)	343/1125	69.5	78/376	79.3
	(ii)	158/483	67.3	38/161	76.4
Chicken vs.	Yeast (i)	282/1121	74.8	43/376	88.6
	(ii)	133/480	72.3	23/161	85.7

Nucleotide and amino acid compositions of actin sequences among different organisms. The number of observed substituted (or unmatched) nucleotides and amino acids in the exon region are shown as total changes for the regions of nucleotide or amino acid sequences compared. Homology values for each comparison are expressed as percentages of the number of matched nucleotides and amino acids contained in the exon, between any two species. Where two values are shown, the top value (i) is the comparison of two entire actin sequences and the bottom number (ii) is the calculation based on comparisons of the carboxy-terminal portion of the actin genes that correspond to the PAc1-A pine DNA sequence determined in this thesis.

TABLE 2: Percentage of Codon Positional Changes

Species	Comparison	%CHANGES AT CODON POSITIONS 1	2	3	PERCENT Total Nucleotide Changes
PAc1-A vs.	SAc1	4.6	2.7	15.4	22.7
	SAc3	3.5	1.4	17.7	22.8
	MAc1	4.8	2.5	17.5	24.8
	Chicken	6.2	5.0	18.6	29.8
	Yeast	9.1	4.3	21.0	34.6
SAc1 vs.	SAc3 (i)	3.8	2.2	16.2	22.6
	(ii)	4.0	1.9	17.4	23.3
	MAc1 (i)	5.3	2.1	18.7	26.1
	(ii)	6.1	2.5	17.1	25.7
	Chicken (i)	5.7	3.6	17.8	27.3
	(ii)	5.6	4.6	18.9	29.1
	Yeast (i)	8.2	3.6	19.8	31.8
	(ii)	8.5	3.7	19.5	31.7
SAc3 vs.	MAc1 (i)	3.7	1.5	17.2	22.6
	(ii)	4.8	1.5	16.6	22.9
	Chicken (i)	4.9	3.7	17.1	25.7
	(ii)	6.0	4.6	19.2	29.8
	Yeast (i)	7.8	3.5	18.5	29.8
	(ii)	8.7	4.1	18.9	31.7
MAc1 vs.	Chicken (i)	5.4	3.5	15.7	24.6
	(ii)	7.5	5.4	15.5	28.4
	Yeast (i)	8.1	3.4	19.0	30.5
	(ii)	8.5	4.6	19.6	32.7
Chicken vs.	Yeast (i)	5.7	2.2	17.1	25.2
	(ii)	5.8	3.3	18.6	27.7

Total number of changes occurring at each position in a codon among different species. Where two values are shown, the top value (i) is the comparison of two entire actins and the bottom number (ii) is the calculation based on comparison of the carboxy-terminal portion of the gene that corresponds to the PAc1-A actin DNA sequence determined in this thesis.

interesting to note that the PAcl-A amino acid sequence is the most homologous to the soybean actin SAc3 gene. It is even more homologous than SAcl. This might be explained by the existence of different sub-classes of actins within an organism that may perform slightly different cellular functions, have different developmental or tissue specificities, or are no longer functional as in the case of pseudogenes (e.g. Little, 1982). For example, many of the human cytoskeletal actins, which have been classified as alpha-actins from the predicted amino acid sequence based on the DNA sequence, appear to be pseudogenes (Ponte *et al.*, 1983). Also, studies in organisms such as *Drosophila* (Fyrberg *et al.*, 1983) and sea urchin (Garcia *et al.*, 1984) have shown that different actins have unique expression patterns dependant on the developmental stage and tissue type analyzed. It is therefore possible that the two soybean genes belong to two different sub-classes and that the SAc3 'type' more closely resembles the PAcl-A 'type' than SAcl.

The comparisons shown in Table 1 rated all amino acid substitions, additions and deletions with equal values. To determine if these calculations gave a true representation of protein evolution, the amino acid substitutions occurring between PAcl-A and SAc3, SAcl, MAcl, chicken, and yeast were assigned values based on the property differences conferred on the actin protein by a given amino acid substitution (Table 3). These values, calculated by Grantham (1974), were based on composition, polarity and molecular volume between the substituted amino acids, where higher values relect greater protein differences. In Table 3, a direct correlation between the crude amino acid homology estimates and the differences determined by Grantham's (1974) parameters was observed. The PAcl-A and SAc3 genes share the lowest amount of amino acid change, whereas the PAcl-A and yeast actin proteins differ the most.

Comparisons among the actin genes of all the species used in the pine actin comparisons were done to determine if comparisons among the carboxy portions of different species were representative of comparisons of entire genes. This was accomplished in two ways, first by comparing entire genes and second, using only the

TABLE 3: Comparisons of Percentage Amino Acid Homology with Weighted Amino Acid Substitutions

Species	Comparison	% A.A. Homology	Weighted A.A. Substitutions
PAc1-A vs.	SAc1	85.1	1376
	SAc3	90.7	854
	MAc1	84.5	1547
	Chicken	82.0	1881
	Yeast	78.3	2020

Amino acid substitutions are weighted according to composition, polarity, and molecular volume with parameters calculated by Grantham (1974). Substitutions resulting in higher values directly correlate with larger protein differences.

carboxy portions of the genes that corresponded to the PAc1-A gene sequence. Similarily, the relative percentage of nucleotide and amino acid changes were compared between the various actin genes when the actin gene is considered as an amino portion or a carboxy portion. These were defined by the regions not encoded or encoded by PAc1-A, respectively (Table 4). While the nucleotide values show little difference, the amino acid differences are consistently greater in the carboxy portion with the exception of the SAc1 and SAc3 actins comparison. Despite this increased variability in the carboxy region, nucleotide and amino acid homology values of all comparisons strongly correspond with evolutionary relatedness of the different species chosen. Thus, results of comparisons of the carboxy terminal portion of actin as represented by the PAc1-A gene seem to correspond to comparisons among entire genes. The plant actins all share higher nucleotide and amino acid homology values than when they are compared with chicken and yeast. The SAc3 and SAc1 actins share amino acid homology values as high as 86.3%, whereas each gene shares 78.9% and 75.2% amino acid homology with yeast, respectively, when their carboxy portions are compared. The chicken-yeast carboxy comparison yielded an amino acid homology value of 85.7%, suprisingly high considering their evolutionary distance. These results suggest that the actin proteins in plants carry out similar specialized functions important for plant cells, and the actin 'type' from chicken codes for a protein necessary for common specialized cellular functions necessary in both yeast and chicken.

D. Localized Variation

It has been noted by others that the first 20 amino acids and the amino acids located between residues 259 and 298 are the two regions of greatest sequence variation among different actins (Vandekerckhove and Weber, 1978b; Vandekerckhove and Weber, 1980). However, Shah *et al.* (1982) have redefined this regional variation based on comparisons of the SAc3 actin sequence to the yeast actin gene sequence. Shah *et al.*

TABLE 4: Species Diversity Comparisons of the Nucleotides and Amino Acids of the Carboxy portion of Actin Proteins

COMPARISON		AMINO PORTION		CARBOXY PORTION	
		%NT	%AA	%NT	%AA
SAc1 VS.	SAc3	22.1	15.7	22.6	13.7
	MAc1	26.4	15.2	25.7	19.9
	Chicken	26.0	16.6	29.1	21.7
	Yeast	31.5	22.6	32.2	24.8
SAc3 vs.	MAc1	22.4	13.0	22.9	14.9
	Chicken	22.6	13.0	29.8	17.4
	Yeast	28.3	18.5	31.7	21.1
MAc1 vs.	Chicken	21.8	10.7	28.4	20.5
	Yeast	28.8	18.6	32.7	23.6
Chicken vs.	Yeast	23.2	9.3	27.7	14.3

The percent differences for nucleotides (%NT) and amino acids (%AA) of species comparisons have been calculated for the amino and carboxy portion of the actin genes. The amino actin portion differences were calculated by subtracting total nucleotide differences detected of the entire gene. The amino portion is defined as actin coding sequences not contained on PAc1-A and the carboxy portion is defined as the region of actin coding sequence contained on PAc1-A.

calculated a value of 37% divergence (or 63% amino acid homology) for the region between amino acid residues 228 to 277 of SAc3 and the corresponding region in yeast (227-275). In order to examine this area of variability in the pine gene the comparisons were subdivided into variable and constant portions. The variable regions were taken as the region encoding amino acids 228 to 277, whereas the constant region was considered to be PAc1-A actin sequence that excludes the above region. These areas were compared to the various species chosen for comparison in this study as shown in Table 5. PAc1-A shares the same region of variability between residues 228 and 277 when compared to the corresponding amino acid sequences of the species in this report. For example, values of only 62.0% and 80.0% amino acid homology were obtained for the variable regions of pine-yeast and pine-soybean (SAc3) comparisons, respectively. These values are considerably lower than the amino acid homology values calculated from the entire 161 amino acid carboxy region (78.3% for pine-yeast and 90.7% for pine-soybean (SAc3) from Table 1). In fact, the values from the variable region separate the actins of different species even more than the comparisons covering the entire carboxy portions of the actin genes. In contrast, homology values were greater for the constant regions when compared to homology values obtained by including the variable regions (Table 1). Similar results were obtained for entire genes when other species were compared. In these comparisons, the variable portions were considered to be the first 20 amino acids in addition to the region between amino acids 228 and 277. The constant regions were calculated by deleting the variable portions from estimates covering entire genes. The same patterns of amino acid differences were observed in homology values for both the 'constant' and 'variable' portions of actins (Table 5) when compared to overall homology values as calculated in Table 1. Yeast actin is the most divergent (or the least homologous) from the other species, chicken is the next most diverged, and the plants are all the least diverged when compared among themselves. Exclusion of the variable region from the pine comparisons increases amino acid homology values in all comparisons. Pine and yeast share 85.6% homology, pine and SAc3 95.5% and, pine and

TABLE 5: Constant vs. Variable Regions of the Actin Gene

		Constant Region		Variable Region	
Species	Comparison	# A.A. Changes	% A.A. Homology	# A.A. Changes	% A.A. Homology
PAc1-A vs.	SAc1	12/111	89.2	12/50	76.0
	SAc3	5/111	95.5	10/50	80.0
	MAc1	13/111	88.3	12/50	76.0
	Chicken	14/111	87.4	15/50	70.0
	Yeast	16/111	85.6	19/50	62.0
SAc1 vs.	SAc3	46/308	85.1	12/70	82.9
	MAc1	50/308	83.8	15/70	78.6
	Chicken	47/308	84.7	24/70	65.7
	Yeast	60/308	80.5	29/70	58.6
SAc3 vs.	MAc1	37/307	87.9	15/70	78.6
	Chicken	34/307	88.9	22/70	68.6
	Yeast	47/307	84.7	27/70	61.4
MAc1 vs.	Chicken	35/306	88.6	21/70	70.0
	Yeast	50/306	83.7	28/70	60.0
Chicken vs.	Yeast	27/305	91.1	16/69	76.8

Amino acid homology of actin genes based on the amino acid composition in the variable and constant regions are defined in the text. Comparisons of species variable regions where pine actin is not involved also include the first 20 amino acids of the amino terminus. In comparisons not involving pine, the constant portion was defined as the entire actin coding sequence excluding the two variable regions.

maize 88.3%. Thus, if this region of variability is excluded in all the species comparisons, the actin gene is even more conserved than the overall estimates in Table 1 indicate. Thus, the regional variability accounts for most if not all of the lower homology values observed in the carboxy portions of actin when compared with the amino portions in Table 4. For example, corn and yeast share 82.0% amino acid homology over this region when the amino acids between codons 228 and 277 were omitted, compared to 82.2% homology observed in the amino portion. The amino acid variability at the amino-terminal end of actin has long been a basis for classification of actin types (i.e. cytoplasmic vs. muscle actins) (e.g. Vandekerckhove and Weber, 1978a). It is possible that the region of variability from residues 228 to 277 in SAc3, PAc1-A and yeast may be a second region that can be used to differentiate actin protein functional differences between species. Alternatively, this may be a region of the actin protein that is not as important to the functional properties of either class of the protein.

E. Codon Positional Changes

Amino acid homology values of the actin protein are consistently higher than nucleotide homology values in all comparisons. This can be explained by the predominance of third codon positional changes that are mostly silent with respect to amino acid replacements (Table 2). This prevalence of third base substitutions has also been observed in comparisons of other actin genes. For example, 79% of all base changes between human and rat actins occur in the third position (Ponte *et al.*, 1984). In addition to third position changes in the various species comparisons (Table 1), many nucleotide substitutions occur in the first position of codons for the amino acids arginine and leucine, which have two alternate coding forms in the first position and in both the first and second position for serine, which has two alternatives in both the first and second position. (Therefore, in all comparisons many of the amino acid changes that

do occur result in changes to functionally similar amino acids). Although synonymous amino acid changes do not affect the protein structure, they could affect gene expression due to RNA sequence changes. For example, secondary structures of RNA may influence transcriptional, processing or translational processes (Lewin, 1985).

F. Codon Useage

The codon useage pattern for PAc1-A (Table 6) resembles that published for SAc3 (Shah *et al.*, 1982). Both genes show a slight bias against codons ending in G or C; 39.6% of PAc1-A codons and 41.8% of SAc3 codons end in G or C. A bias for codons specifying the amino acids leucine, cysteine, arginine and glycine exists in pine. Only three of a possible six codons for both leucine and arginine are used. Leucine is coded only by CTT, CTC, and TTG codons and arginine by AGG, AGA, and CGT Cysteine is only coded by TGT and glycine is coded by GGT and GGA most of the time (six and four times, respectively), while GGC is used only once and GGG not at all. However, it should be stressed that these biases may be typical of only the carboxy portion of the pine actin and not be truly representative of the entire gene or general codon useage in pine. SAc3 shows a bias of codon utilization for leucine, isoleucine, serine, proline, threonine, alanine and arginine (Shah *et al.*, 1982). In these cases, with the exception of isoleucine, all codons of the general form XYG are not represented. The TTA form of leucine is also not present and the ATA codon (isoleucine) is expressed only once out of a possible 28 times. The two other plant actin genes, SAc1 and MAc1, also display a bias in codon utilization of the amino acids valine, serine, leucine and isoleucine. In these cases, codons of the general form XYA, are used less than expected (Shah *et al.*, 1983). Thus, although all four plant actin genes share a bias against certain leucine codons, the PAc1-A gene, MAc1, and SAc1 genes share no other similarities in their codon useage patterns. The SAc3 gene is similar to the PAc1-A gene because both genes share a slight bias of codons with the general form

TABLE 6: Codon Useage in the Carboxy terminus of a Lodgepole Pine Actin Gene

Phe	TTT	3	Ser	TCT	5	Tyr	TAT	5	Cys	TGT	3
	TTC	2		TCC	2		TAC	2		TGC	0
Leu	TTA	0		TCA	3	Ala	GCT	5	Glu	GAA	4
	TTG	2		TCG	1		GCC	2		GAG	10
	CTT	7		AGT	5		GCA	4	Trp	TGG	2
	CTC	2		AGC	3		GCG	1	Arg	CGT	2
	CTA	0	Pro	CCT	3	His	CAT	1		CGC	0
	CTG	0		CCC	3		CAC	0		CGA	0
Ile	ATT	5		CCA	1	Gln	CAA	2		CGG	0
	ATC	6		CCG	0		CAG	4		AGA	3
	ATA	1	Thr	ACT	2	Asn	AAT	2		AGG	1
Met	ATG	7		ACC	3		AAC	1	Lys	AAA	5
Val	GTT	5		ACA	3	Gly	GGT	6		AAG	6
	GTC	1		ACG	1		GGC	1	Stop	TAA	1
	GTA	2	Asp	GAT	5		GGA	4		TAG	0
	GTG	1		GAC	1		GGG	0		TGA	0

Total number of occurrences of each codon in the PAcl-A actin coding sequence.

XYG. On the other hand, the SAc3, SAcl, and MAcl genes share codon utilization biases; those for leucine, isoleucine, and serine. Therefore, the SAc3 gene appears to bridge the gap observed between PAcl-A and SAc3 and MAcl when codon useage patterns are analyzed.

G. Intron

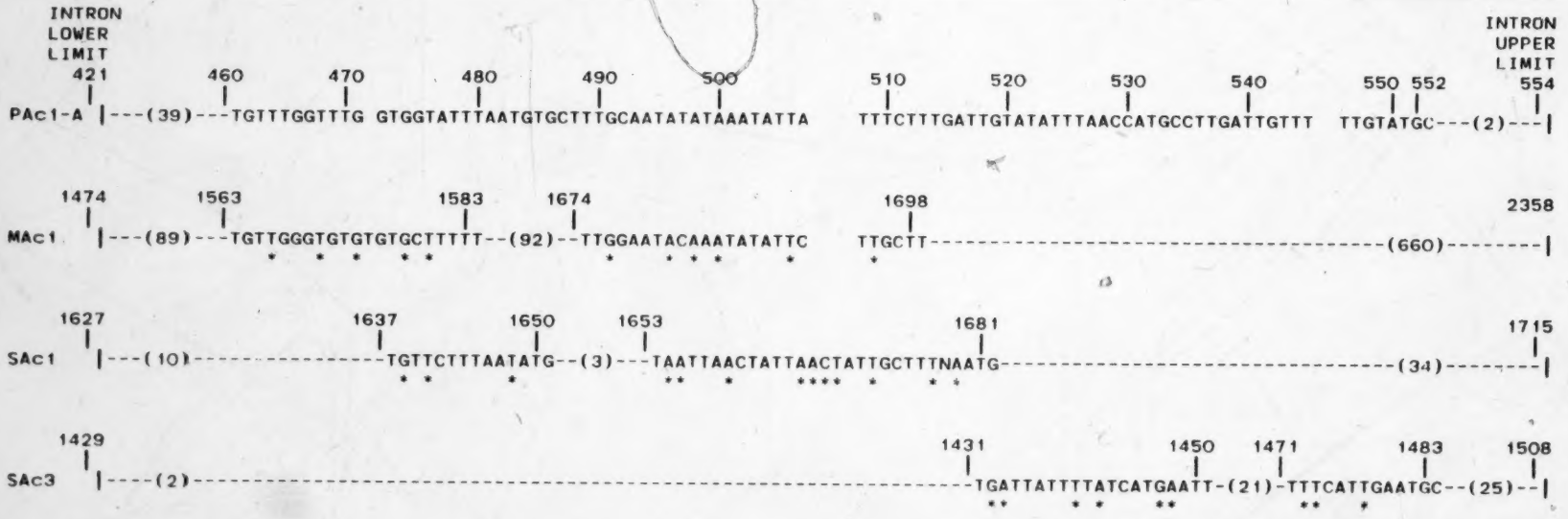
Previous comparisons between MAcl and two soybean actins (SAcl and SAc3) have shown that all three plant actins share identical intron positions (Shah *et al.*, 1983). This is in contrast to the lack of intron position conservation between and within animal species (Fyrberg *et al.*, 1981). It has been suggested by Shah *et al.* (1983) that positional intron conservation in the maize and soybean actins may indicate that plant actin genes have evolved from a common ancestral actin gene containing a minimum of three introns. Based on this hypothesis, it was suggested that positions of introns between divergent plant species may also be conserved. The PAcl-A intron after codon position 355 corresponds to the third position introns found in both soybean and maize and supports this intron position conservation view. Unfortunately, the other two introns found in both maize and soybean occur at positions that are not present on the cloned PAcl-A fragment.

The nucleotides found at the exon/intron junctions in PAcl-A follow the consensus splice site junctions (Breathnach *et al.*, 1978). The intron donor nucleotides are GGT and the acceptor nucleotides are CAG, which correspond to most of the animal and all of the plant actin genes reported. The endpoints of the intron within the donor and acceptor nucleotides for pine are:

...Exon...AG/GT...Intron...CAG/AT...Exon.

Partial homologies within the third intron among the four different plant actin genes (PAcl, SAcl, SAc3 and, MAcl) were noted. These homologies covered short regions of the intron. Figure 11 shows a possible alignment achieving homology values

Figure 11: Comparison of Homology in the Third Intron of the Plant Actins.



Partially homologous sequences among the plant actins PAc1-A, MAc1, SAc1, and SAc3..
Nucleotides with asterisks (*) underneath indicate nucleotide differences between the species intron where they occur and the PAc1-A intron.
Numbers in brackets () show the numbers of unmatched nucleotides occurring at various locations within the intron required to maintain the intron sizes.

University of Alberta

of 76.1%, 70.0% and, 72.7% when the PAc1 intron sequence was compared with the third introns of MAc1, SAc1 and, SAc3, respectively. These data would appear to support the hypothesis that the plant actin genes evolved from a common ancestral gene. In addition, the retention of partial intron homologies could reflect that the introns of these plant species are processed by similar mechanisms.

IV. CONCLUSION

The nucleotide and amino acid sequence of the carboxy terminal portion of a lodgepole pine actin gene, PAc1-A, has been presented and compared to published actin sequences of maize (MAc1), soybean (SAc3 and SAc1), chicken and yeast. This 750 nucleotide portion contains coding sequences for amino acids 216 to 376, as defined by the SAc3 sequence, and is interrupted by an intron 130 nucleotides in length.

The actin gene was defined by its carboxy portion and amino portion where these two regions were defined by sequences contained on the PAc1-A fragment and not so contained, respectively. Nucleotide and amino acid sequences of entire actin coding sequences of maize, soybean, chicken, and yeast were compared against portions of the carboxy-terminal portions of the actin genes represented by the PAc1-A fragment. With the exception of the SAc1-SAc3 comparison, the carboxy portion of the various actin genes displayed consistently less homology among other organisms than did its amino portion. This disparity was explained by the existence of a highly variable region located between codons 228 and 277, regardless of the two different actin sequences compared. When this region was disregarded, little or no discrepancy in percentage amino acid homology existed between the amino and carboxy portions of any two genes. Despite the overall increase in homology when the variable regions were disregarded, the overall trends of percent amino acid homology between different actin genes were identical to the comparison of the whole genes regardless of whether this variable region was included or not in the respective comparisons. Comparisons among the variable regions of the different organisms resulted in greater observed differences between any two actin genes with the exception of the two soybean clones, SAc1-SAc3. Thus, it was concluded that comparisons of actin gene portions were representative of comparisons between entire genes.

In general, amino acid homology was consistently high in all comparisons despite large evolutionary distances. This conservation of the actin gene likely reflects the importance of the protein in carrying out basic cellular functions. The nucleotide

homology values are lower than the amino acid homology, which reflects that most nucleotide substitutions resulted in 'silent' amino acid changes. Morever, a high percentage of similar amino acid changes that did occur resulted in relatively innocuous differences between different proteins. However, trends were observed within the different species comparisons to separate the actins into distinct classes, possibly based on slightly different functional differences among proteins. When the plant actins were compared with each other, higher nucleotide and amino acid values were observed than when any plant actin was compared to either yeast or chicken. In addition, yeast homology values were much lower in comparisons to plants than in the comparison to chicken. It is possible that at least two classes of actins have been compared. That is, the plant actins may all belong to one class, which requires the actin protein to carry out highly specialized functions in plant cells, while the yeast and chicken actins used for comparison in this thesis appear to belong to a second class, where the function of the actin protein may be similar. It is also possible that subclasses exist within these classes. PAc1-A and SAc3 actins appear to share the highest protein homology, which may be related to similar specialized protein functions that the two particular actins must perform. In fact, the constant regions of the SAc3 actin appear to be more closely related to the constant regions of the other species actins than SAc1.

The intron found in PAc1-A occurs at the same position as the third intron found in SAc3, SAc1 and MAc1. The first and second introns occur after amino acids 19 and 150, respectively, which are not contained on the PAc1-A fragment. The positional intron conservation displayed between the above mentioned species is an unusual feature when considering the lack of positional intron conservation displayed between other actin genes within and between different species. Shah *et al.*, (1983) suggested that this feature may reflect the evolution of plant genes from a common ancestral actin gene containing a minimum of three introns. The possibility also exists that strong selective pressure may have existed such that only introns in some locations were tolerated in the plant actin genes.

Although resolution of individual bands in the Southern data presented is low, a minimum of eight and nine bands can be observed in *EcoRI* and *HindIII* digested pine genomic DNA hybridized with PAc1. Thus, it seems likely that the pine multigene family consists of a minimum of eight actin genes.

There are two logical followups of this project. Firstly, isolation of the entire lodgepole pine actin gene, either from a cDNA or genomic DNA library, would yield more information about possible functions or evolutionary similarities and differences between pine and other species actin proteins. Isolation of the entire PAc1-A actin from the pine actin multigene family would be facilitated by the use of a specific probe derived from PAc1-A. Specifically, the intron sequence or 3'-flanking sequences could be used to probe a library, since little conservation of these actin sequences has been detected either within or between different species. The second study would investigate the expression and control of expression of the entire cloned pine actin gene. Determination of whether the pine actin gene is transcribed could be achieved by hybridization of PAc1-A to Northern blots containing polyadenylated mRNA.

This study has shown that available molecular technology can be applied towards the isolation and characterization of genes of lodgepole pine. Although isolation of the pine actin carboxy-terminus represents basic research, it is conceivable that the techniques and modifications developed in this thesis could have more applied implications in the long term. For example, traits such as insect and disease resistance or traits affecting growth components of trees may possibly be transferred within or among species.

V. LITERATURE CITED

Barton, K.A., Binns, A.N., Matzke, A.J.M., and Chilton, M. 1983. Cell 32: 1033-1043.

Barton, K.A., and Brill, W. 1983. Science 219: 671-676.

Bergsma, D.J., Chang, K.S., and Schwartz, R.J. 1985. Mol. Cell Biol. 5: 1151-1162.

Bevan, M., and Chilton, M.D. 1982. Ann. Rev. Genet. 16: 357-384.

Birnstiel, M., Busslinger, M., and Strub,K. 1985. Cell: 349-359.

Breathnach, R., Benoist, C., O'Hare, K., Gannon, F., and Chambon, P. 1978. Proc. Natl. Acad. Sci, 75: 4853-4857.

Brown, S., Levinson W. and, Spudich, J.A. 1976. J. Supramol. Structure 5: 119-130.

Chilton, M. 1983. Biotechnology. April 163-164.

Cocking, E.C.,Davey, M.R.,Pental, D.,and Power, J.B. 1981. Nature 293: 265-270.

Coen, E.S., Thoday, J.M., and Dover, G. 1982. Nature 295: 564-568.

Condeelis, J.S. 1974. Exp. Cell Res. 88: 435-439.

Elzinga, M. and Lu, R.C. 1976. In Contractile Systems in Non-Muscle Tissues, S.V. Perry, A. Margreth and R.A. Adelstein, eds. (Amsterdam) pp. 29-37.

Engel, J.N., Gunning, P.W., and Kedes, L. 1981. Proc. Natl. Acad. Sci. 78: 4674-4678.

Faulkner, R. 1975. Seed Orchards. Forestry Commission Bulleton. Faulkner, R. ed., Forestry Commission, 54

Feinberg, A.P., and Vogelstein, B. 1983. Analytical Biochem. 132: 6-13.

Feinberg, A.P., and Vogelstein, B. 1984. Analytical Biochem. 137: 266-267.

Files, J.G., Carr, S., and Hirsh, D. 1983. J. Mol. Biol. 164: 355-375.

Firtel, R.A. 1981. Cell 24: 6-7.

Forer, A., and Jackson, W.T. 1975. Cytobiologie 10: 217-226.

Fowler, V. and Bennett, V. 1978. J. Supramol. Struct. 8: 215-221.

Fyrberg, E., Bond,B., Hershey, N.D., Mixter, K.S., and Davidson, N. 1981. Cell 24: 107-116.

Fyrberg, F.A., Mahaffey, J.W., Bond, B.J., and Davidson, N. 1983. Cell 33: 115-123.

Gallwitz, D. and Seidel, R. 1980. Nucleic Acids Res. 8: 1043-1059.

Gallwitz, D. and Sures, I. 1980. Proc. Natl. Acad. Sci. USA Vol 77: 2546-2550.

Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B., and Chambon, P. 1979. Nature 278: 428-434.

Garcia, R., Paz-Aliaga, B., Ernst, S.G., and Crain, Jr., W.R. 1984. Mol. Cell. Biol. 4: 840-845.

Gawadi, N. 1971. Nature 234: 410.

Gilbert, W. 1978. Nature 271: 501.

Glenney, Jr. J.R., and Glenney, P. 1984. Cell 37: 743-751.

Glover, D.M. 1980. Genetic Engineering. Chapman and Hall Ltd., Methuen Inc., New York.,eds.

Godson, G.N. 1980. Fed. Proc. 39: 2822-2829.

Goodman, S.R. and Branton, D. 1978. J. Supramol. Struct. 8: 455-463.

Grantham, R., 1974. Science 185: 862-864.

Gruenstein, M., and Hogness, D.S. 1975. Proc. Natl. Acad. Sci. USA 72: 3961-3965.

Gruenstein, M., Rich, A., and Weihing, R.R. 1975. J. Cell Biol. 64: 223-234.

Hartwig, J.H., and Stossel, T.P. 1975. J. Biol. Chem. 250: 5696-5705.

Hentschel, C., Irminger, J.C., Bucher, P. and Birnstiel, M.L. 1980. Nature 285: 147-151.

Izant, J.G., and Lazarides, E. 1974. Proc. Natl. Acad. Sci. USA. 74: 1450-1454.

Kahn, M., Kolter, R., Thomas, C., Figurski, D., Meyer, R., Remaut, E., and Helinski, D.N. 1979. Methods in Enzymology 68: 268-280.

Kaine, B.P., and Spear, B.B. 1982. Nature 295: 430-432.

Katcoff et al., 1980. Proc. Natl. Acad. Sci. 77: 906-964.

Kersey, Y.M., and Wessells, N.K. 1976. J. Cell Biol. 68: 264-275.

Kindle, K., and Firtel, R.A. 1978. Cell 15: 763-778.

Klessig, D.F., and Berry, J.O. 1983. Plant Molecular Biology Reporter 1: 12-16.

Koch, G.L.E., and Smith, M.J. 1978. Nature 273: 274-278.

Korn, E.D. 1976. Trends Biochem. Sci. 1: 55-58.

Korn, E.D. 1978. Proc. Natl. Acad. Sci. USA 75: 588-599.

Kost, T.A., Thiodorakis, N., and Hughes, S.H. 1983. Nuc. Acids Res. 11: 8287-8301.

Lewin, B. Genes II. 1985. Cell, ed., John Wiley and Sons, New York.

Little, P.F.R. 1982. Cell 28: 683-684.

Maniatis, T., Fritsch, E.F., and Sambrook, J. Molecular Cloning: A laboratory manual. 1982. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

McKeown, M., and Firtel, R.A. 1981. Cell 24: 799-807.

McKeown, M., Taylor, W.C., Kindle, K.L., and Firtel, R.L. 1978. Cell 15: 789-800.

Messing, J. Methods in Enzomology. 1983. (Wu, R., Grossman, L., and Moldave, K., eds.) Vol. 101: 20-78.

Miksche, J.P., and Dhillon, S.S. 1981. Nucleic acid research techniques in forest genetics IN Advances in Forest Genetics, P.K. Khosla. (Ed.) Ambika, New Delhi, India.

Minty, A.J., Alonso, S., Guenet, J.L., and Buckingham, M.E. 1983. J. Mol. Cell. Biol. 167: 77-101.

Nagao, R.T., Shah, D.M., Eckenrode, V.K., and Meagher, R.B. 1981. DNA 1: 1-9.

Nakajima-Iijima, S., Hamada, H., Reddy, P., and Kakunaga, T. 1985. Proc. Natl. Acad. Sci. USA 82: 6133-6137.

Netzer, W.J. 1984. Bio/Technology 2: 596.

Ng, R., and Abelson, J. 1980. Proc. Natl. Acad. Sci. USA 77: 3912-3916.

Ng, R., Domdy, H., Larson, G., Rossi, J.J., and Abelson, J. 1985. Nature 314: 183-184.

Norgard, M.V., Keem, K., and Monaham, J.J. 1978. Gene 3: 279-292.

Norrander, J., Kemper, T., and Messing, T. 1983. Gene 26: 101-106.

Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z., and Yaffe, D. 1983. Nucl. Acids Res. 11: 1759-1771.

Pesacreta, T.C., Carley, W.W., and Parthasarathy, M.V. 1982. Proc. Natl. Acad. Sci. USA 79: 2898-2901.

Pollard, T., and Weihing, R. 1974. In Critical Reviews in Biochen., 2, G. Fasman, ed. (Cleveland, Ohio:CRC Press).

Ponte, P., Gunning, P., Blau, H., and Kedes, L. 1983. Mol. Cell Biol. 3: 1783-1791.

Ponte, P., Ng, S., Engel, J., Gunning, P., and Kedes, L. 1984. Nuc. Acids Res. 12: 1687-1696.

Pribnow, D. 1975. Proc. Natl. Acad. Sci. USA 72: 784-788.

Proudfoot, N.J., and Brownlee, G.G. 1976. Nature 263: 211-214.

Sanger, F., and Coulson, A.R. 1978. FEBS Lett. 87: 107

Sanger, F., Coulson, A.R., Hong, G.F., Hill, D.F., and Peterson, G.B. 1982. J. Mol. Biol. 162: 729-773.

Sanger, F., Nicklen, S., and Coulson, A.R. 1977. Proc. Natl. Acad. Sci. 74: 5463-5467.

Schell, J. and VanMontagu, M. 1983. Biotechnology 1: 175-180.

Schleif, R.F., Wensink, P.C. 1981. Practical Methods in Molecular Biology. Springer-Verlag, New York, Heidelberg Berlin.

Seifriz, W. 1942. Am. Soc. Plant Physiol. Ames, IA.

Shah, D.M., Hightower, R.C., and Meagher, R.B. 1982. Proc. Natl. Acad. Sci. 79: 1022-1026.

Shah, D.M., Hightower, R.C., and Meagher, R.B. 1983. J. Mol. Appl. Gen. 2: 111-126.

Shepard, J., Bidney, D., Barsby, T., and Kemble, R. 1983. Science 219: 683.

Scheer, U., Hinssen, H., Franke, W.W., and Jockusch, B.M. 1984. Cell 39: 111-122.

Schuler, M.A., McOsker, P., and Keller, E.B. 1983. Mol. Cell. Biol. 3: 448-456.

Stadem, R. 1977. Nuc. Acids Res. 4: 4037-4051.

Stadem, R. 1978a. Nuc. Acids Res. 5: 1013-1015.

Stadem, R. 1978b. Nuc. Acids Res. 6: 2601-2609.

Stadem, R. 1980a. Nuc. Acids Res. 8: 817-825.

Stadem, R. 1980b. Nuc. Acids Res. 8: 3673-3694.

Tilney, L.G. 1976. J. Cell Biol. 69: 73-89.

Tilney, L.G., and Detmers, P. 1975. J. Cell Biol. 66: 508-520.

Tobin, S.L., Zulauf, E., Sanchez, F., Craig, E.A., and McCarthy, B.J. 1980. Cell 19: 121-131.

Vandekerckhove, J., and Weber, K. 1978a. Nature 276: 720-721.

Vandekerckhove, J., and Weber, K. 1978b. J. Mol. Biol. 126: 783-802.

Vandekerckhove, J., and Weber, K. 1980. Nature 284: 475-477.

Waters, S.H., Distel, R.J., and Hecht, N.B. 1985. Mol. Cell. Biol. 5: 1649-1654.

Watson, B., Currier, T., Gordon, M., Chilton, M.D., and Nester, E. 1975. J. Bacteriol. 123: 255-264.

Wegner, A. 1985. Nature 313: 97-98.

Weihing, R.R. 1976. Cell Biology, Altman P.C. and Dittmer, eds. D.S.

Williamson, R.E. 1980. Can. J. Bot. 58: 766-772.

Yanisch-Perron, C., Vieira, J., and Messing, J. 1985. Gene 33: 103-119.

Zafar, R.S., and Sodja, A. 1983. Biochem. Biophy. Res. Comm. 111: 67-73.

Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D., and Nudel, U. 1982. Nature 298: 857-859.

Zimmer, E.A., Rivin, C.J., and Walbot, V. 1981. Plant Mol. Biol. Newsletter 2: 93-96.